CIP Competitiveness and Innovation Framework Programme
CIP-Pilot Actions, 2007-2013
CIP-ICT-PSP-2012-6

Project **CIP-Pilot 325101 / OpenScienceLink**
Deliverable **D3.2**
Distribution **Public**

**http://OpenScienceLink.eu**

# Legal and IPR Management Framework Specification

Sara Hugelier, Marie-Christine Janssens, Freyja van den Boom and Griet Verhenneman

Status: **Revision 2**
April 2015

**Project**

| | |
|---|---|
| Project ref.no. | CIP-Pilot 325101 |
| Project acronym | OpenScienceLink |
| Project full title | Open Semantically-enabled, Social-aware Access to Scientific Data |
| Project site | http://projects.biotec.tu-dresden.de/OpenScienceLink / |
| Project start | February 2013 |
| Project duration | 36 months |
| EC Project Officer | David Guedj |

**Deliverable**

| | |
|---|---|
| Deliverable type | Report |
| Distribution level | Public |
| Deliverable Number | D3.2 revision 2 |
| Deliverable title | Report Title |
| Contractual date of delivery | M25 (March 2015) |
| Actual date of delivery | April 2015 |
| Relevant Task(s) | WP3/Task3.5 |
| Partner Responsible | KU Leuven |
| Other contributors | |
| Number of pages | |
| Author(s) | Sara Hugelier, Marie-Christine Janssens, Freyja van den Boom and Griet Verhenneman |
| Internal Reviewers | TUD |
| Status & version | Revision 2 |
| Keywords | Legal framework, Intellectual property rights, data protection, privacy, licensing framework, public sector information directive |

# Abstract

The objective of this deliverable is to clarify the legal state of the art and rules applicable to the OpenScienceLink project and to ensure that the development and workings of the OpenScienceLink platform are compatible with the different regulatory frameworks.

This deliverable will provide an updated legal analysis of the applicable legal framework to the OpenScienceLink project. The legal analysis focuses on the explanation of the most important legal concepts relevant for the development of the platform.

(a) The first part will introduce the international and EU policy on open access, including relevant such as the PSI Directive dealing with the re-use of public sector information.
(b) The second part of this deliverable provides a legal evaluation of the privacy issues associated with the acquisition and the privacy issues for researchers.
(c) The third part will focus on the analysis of intellectual property right issues. In particular, what are the implications of copyright law and the Database Directive on research data made available on the platform.
(d) The fourth part of this deliverable will focus on setting up a licensing framework.
(e) The final part will provide information on the issue of Internet service providers' liability under the E-Commerce Directive.

This deliverable should provide a clear description and explanation of the applicable legal framework to the OpenScienceLink platform, and more in general, the legal aspects of open access to research data. It should be perceived as an introductory document.

More detailed explanations of selected legal issues will be subject to separate analysis

# Contents

# 1 Introduction

The main goal of the OpenScienceLink project is to introduce and pilot a holistic approach to the publication, sharing, linking, review and evaluation of research results, based on open access to scientific information. The project's holistic approach will empower a novel ecosystem for open access, which will provide a range of added-value services for all stakeholders.

The OpenScienceLink project aims to provide a universal well-structured repository of scientific and research data –currently focusing on biomedical and clinical research- for experimentation and benchmarking of pertinent research works in a given thematic area. Moreover, the OpenScienceLink web-based platform will enable/provide:

- Creation, establishment, maintenance and evaluation of open data journals;
- Publishing and sharing of publications and experimental datasets, as well as their linking with researchers and scholars;
- Tools and techniques for structured sharing and exchange of experiences across all the stakeholders of the open access paradigm and associated value chains (including publishers, researchers and scholars);
- Novel tools for reviewing articles, aiding editors to find competent reviewers, and referees to gain instant access to all supporting information (such as relevant research articles and datasets) and
- Introduction and management of new metrics of scientific performance for individual scholars, but also groups, clusters and communities of scholars.

In this document, we cover the main requirements based on the European legal framework that the OpenScienceLink project needs to comply with to ensure the legality of the project.

The following will be addressed in the subsequent sections:

- European policies on open access to scientific information. We have included the Directive on the re-use of public sector information
- Data protection regulations and more specifically, The European data protection Directive and reform.
- The intellectual property legislations applicable to the content and services of the platform. We will cover copyright protection and the proposed reform and the database directive.
- The licensing framework. In this part we present an overview of the different licensing options based on the requirements of the previous sections
- The liability exemption for of Internet service providers. We will look at the requirements under the E-Commerce Directive.

# 2 Open access to scientific information

## 2.1 Introduction

Over the last few years programs such as FP7 and in particular Horizon 2020 have invested in the development of a true open-access-ecosystem. [1] The rapid development in computing technology and the Internet has made new applications to maximize the dissemination and impact of new research possible. Not only research data but also government data (like statistical data) continuously become available via repositories and portals for commercial and non-commercial use

Having access to research data is playing an ever more important role.[2] The OpenScienceLink holistic approach to the publication, sharing, linking, review and evaluation of research results aims to contribute to the Open Access ecosystem for scientific information. In this chapter we will give an overview of the Open Access policies, recommendations and principles which are relevant for the purpose of the OpenScienceLink project.[3]

## 2.2 Open Access developments

Open Access is a means of disseminating scholarly research that breaks from the traditional subscription model of academic publishing.[4] A vast amount of declarations, guidelines, and statements have been published on the subject of Open Access. Early examples include the Budapest Open Access Initiative, [3] the Bethesda Statement on Open Access Publishing, [4] and the Berlin Declaration, [5]

Together these documents provide what is required for a work to qualify as 'Open Access'

- The author(s) and right holder(s) grant(s) to all users a free, irrevocable, worldwide, right of access, and a license, to copy, use, distribute, transmit and display the work publicly and to make and distribute derivative works, in any digital medium for any responsible purpose, subject to proper attribution of authorship, as well as the right to make small numbers of printed copies for their personal use.
- A complete version of the work and all supplemental materials, including a copy of the permission as stated above, in an appropriate standard electronic format is deposited (and thus

[1] Recital 28 and Article 18 of Regulation (EU) No .../2013 of the European Parliament and of the Council establishing Horizon 2020 - the Framework Programme for Research and Innovation (2014 2020) and repealing Decision No 1982/2006/EC. In addition, the Commission has set up a pilot scheme on open access to, and re-use of research data in particular generated by projects in selected areas of Horizon 2020.

[2] *Knowledge and innovation are the true currency of our modern* economy European Commission, A reinforced European Research Area Partnership, COM (2012) 392.

[3] For example: the Recode project, OpenAIRE and OpenAirePlus which links publications with datasets and the OpenScienceLink project. In addition, the Commission has set up a pilot scheme on open access to, and re-use of research data in particular generated by projects in selected areas of Horizon 2020.

[4] Hugelier S. and Janssen K. (2013), Open Data as the standard for Europe? A critical analysis of the European Commission's proposal to amend the PSI Directive, to be published in EJLT, (1) 24.

published) in at least one online repository using suitable technical standards (such as the Open Archive definitions) that is supported and maintained by an academic institution, scholarly society, government agency, or other well-established organization that seeks to enable open access, unrestricted distribution, inter-operability, and long-term archiving.

### 2.2.1 OECD principles and guidelines

The Organization for Economic Co-operation and Development (OECD) provided recommendations to help promote and improve access to publicly funded research data. They published a declaration with a set of principles and guidelines to help governments set up an open access policy.[5]

- Open access to research data from public funding should be *easy, timely, user-friendly and preferably Internet-based;*
- Information on research data and data-producing organisations, documentation on the data and specifications of conditions attached to the use of these data should be internationally available in a transparent way, ideally through the internet;
- Data access arrangements should *respect the legal rights and legitimate interests* of all stakeholders in the public research enterprises;
- Data access arrangements should *consider the applicability of copyright or of other intellectual property laws* that may be relevant to publicly funded research databases;
- Institutional arrangements for the management of research data should be based on the *relevant professional standards and values* embodied in the codes of conduct of the scientific communities involved;
- Research data must comply with explicit quality standards;
- Specific attention should be devoted to developing high-level security techniques to *guarantee the security and integrity of research data;*
- Due consideration should be given to the *sustainability of access* to publicly funded research data as a key element of the research infrastructure.[6]

### 2.2.2 European policy on Open Access

In order to improve access to scientific information and to boost the benefits of public investment in research, the European Commission has implemented Open Access into its Horizon 2020 programme. It has been supporting a number of initiatives and projects with regards to open access to scientific information and research data. [7]

More recently the European Commission set out a Communication towards better access to scientific information. [8] It stated that 'among the actions to be taken under the 'Digital Agenda', publicly funded research should be widely disseminated through open access publication of scientific data and papers'.

---

[5] OECD Principles and Guidelines for Access to Research Data from Public Funding (2006). Retrieved from http://www.oecd.org/sti/sci-tech/38500813.pdf -Increases the returns from public investment in this area; Reinforces open scientific inquiry; Encourages diversity of studies and opinion; Promotes new areas of work; and Enables the exploration of topics not envisioned by the initial investigators.

[6] *Ibid.*

[7] Europe 2020. Retrieved from http://ec.europa.eu/europe2020/index_en.htm relevant projects include OpenAIRE, the RECODE project and FOSTER OPEN SCIENCE.

[8] Communication from the Commission towards better access to scientific information, COM (2012) 401 final (1)

This communications fits in the Europe 2020 Strategy which underlines the central role of knowledge and innovation in generating growth.[9]

By setting out these measures and ensuring that EU-funded research becomes widely and publicly accessible, the EU intends to lead by example, inspiring other governments and institutions to do the same. [10]

#### 2.2.2.1 PSI –Directive

The Directive on the re-use of public sector information ('PSI Directive') entered into force on 31 December 2003. The PSI-directive falls under the EU policy on open data, focussing on generating value through re-use of a specific type of data – public sector information.[11]

The main purpose of the PSI Directive is to establish' a minimum set of rules governing the re-use and the practical means of facilitating reuse of existing documents held by public sector bodies of the Member States'. [12] While the importance of PSI for the entire society was recognized, the main target group of the PSI Directive was the information industry for creating information products and services based on data from the public sector.[13] The PSI Directive focuses on the re-use of information rather than on the access of citizens to information. However the recent studies provide useful information on how to make re-use possible. This is also useful for open access policies in general.

*Licenses for re-use*

The directive states that if licences are used they *'shall not unnecessarily restrict possibilities for re-use and shall not be used to restrict competition'*.[14] With respect to the licensing framework the following recommendations were made which are relevant for OpenScienceLink: [15]

[9] Communication from the Commission towards better access to scientific information, COM (2012) 401 final (1). Essential is that the document sets out a number of key measures to improve the access to and preservation of scientific information on a vast number of areas; policy measures; EU-funded research, funding for infrastructures and projects and coordination beyond the EU.

[10] See COMMISSION RECOMMENDATION of 17 July 2012 on access to, and preservation of scientific information (2012/417/EU).

[11] That is, all the information that public bodies produce, collect or pay for, sometimes referred to as 'government data'. Examples are: geographical information, statistics, weather data, data from publicly funded research projects, and digitised books from libraries. Retrieved from http://ec.europa.eu/digital-agenda/en/european-legislation-reuse-public-sector-information#revision-of-the-directive .

[12] Recital (8) Public sector bodies collect, produce, reproduce and disseminate documents to fulfil their public tasks. Use of such documents for other reasons constitutes a re-use.

[13] In case OpenScienceLink wants to re-use public sector research data, it will have to be assessed whether the institution holding the data falls under the PSI Directive or whether this institution falls under one of the exemptions, for example, when this concerns a publicly funded research institution.

[14] Directive 2003/98/EC of the European parliament and of the council of 17 November 2003 on the re-use of public sector information.

[15] Recital (17) in some cases the re-use of documents will take place without a licence being agreed upon. In other cases a licence will be issued imposing conditions on the re-use by the licensee dealing with issues such as liability, the proper use of documents, guaranteeing non-alteration and the acknowledgement of source. If public sector bodies license documents for re-use, the licence conditions should be fair and transparent. Standard licences that

- The chosen license should be both standard and re-usable, i.e. contain terms that remain fixed both for the licensor and the licensee, follow a formal and open upgrade process, that there is a community working on their upgrade and that they are global in scope.
- Refrain from licensing under licenses that reserve commercial rights. Licenses that contain a Non-commercial clause will forbid a large number of uses that should be available to users, (not to mention the risk inherent in conflicting interpretations in the Non-commercial term itself).
- Avoid using licenses containing a copy left/Share Alike clause, which requires users to release any derivative works under the same license. These licenses raise interoperability issues and are difficult to enforce if the re-user wants to aggregate or mine a large amount of other datasets submitted under different licenses.

*Note on reform*

On 13 June 2013, the European Parliament formally adopted the updated EU rules on the re-use of public sector information.[16] The review of the Directive is one of the key actions of the "Digital Agenda for Europe". The new 2013-PSI Directive has to be transposed in the Member States by July 2015. As a consequence, PSI legislation may vary substantially depending on which Member State the public sector body is established and whether the Member State has already transposed the new PSI Directive into national legislation.

#### 2.2.2.2 Horizon 2020

Open Access encompasses a range of components such as readership, reuse, copyright, posting, and machine readability[17]. Within these areas, publishers and funding agencies have adopted many different policies, some of which are more open and some less open.[18]

In principle " Open scientific research data should be easily discoverable, accessible, assessable, intelligible, useable, and wherever possible interoperable to specific quality standards"[19].

The European commission in its guidelines states that *Open access to research data refers to the right to access and re-use digital research data. Openly accessible research data can typically be accessed, mined, exploited, reproduced and disseminated free of charge for the user.*[20]

The European commission in its guidelines states that 'scientific information' can refer to

---

are available online may also play an important role in this respect. Therefore Member States should provide for the availability of standard licences.

[16] Directive 2013/37/EU of the European parliament and of the council of 26 June 2013 amending Directive 2003/98/EC on the re-use of public sector information, *OJ*. L175/1.

[17] In general it is understood to refer to *information, in particular facts or numbers, collected to be examined and considered as a basis for reasoning, discussion or calculation*. The final report can be accessed here: https://ec.europa.eu/digital-agenda/sites/digital-agenda/files/Report_2013-07-OpenResearchData-Consultation-FINAL1.pdf.

[18] Laakso, M. (2014). Green open access policies of scholarly journal publishers: a study of what, when, and where self-archiving is allowed. Scientometrics. In press. http://dx.doi.org/10.1007/s11192-013-1205-3

[19] The G8 Science Ministers Statement https://www.gov.uk/government/news/g8-science-ministers-statement.

[20] Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020 Version 1.0 11 December 2013.

1. peer-reviewed scientific research articles (published in scholarly journals) [21]
2. Research data (data underlying publications, curated data and/or raw data)

It further defines 'Research data' as

[*I*]*information, in particular facts or numbers, collected to be examined and considered and as a basis for reasoning, discussion, or calculation.*[16] Examples of Research data include statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings and images.

Under Horizon 2020, the requirements for open access to all peer-reviewed scientific publications are [22]

> Step 1 beneficiaries must deposit a machine-readable electronic copy[23] of the published version or final peer-reviewed manuscript accepted for publication in a repository[24] for scientific publications.
>
> Step 2: after depositing publications and, where possible, underlying data, beneficiaries must ensure open access to the deposited publication via the chosen repository.

- Beneficiaries must also ensure open access — via the repository — to the bibliographic metadata that identify the deposited publication. The purpose of the requirement on metadata is to maximise the discoverability of publications. [25]

The challenge for the OpenScienceLink project and Open Access policies in general is how to find the right balance of 'openness' based on the requirements between the making works freely available with as little restraints as possible without infringing upon rights of any of the stakeholders involved. [26]

---

[21] An example of an OA Policy is that of Nature and the Nature research journals, who support the 'green' open access route: Their policy reads as follows: *papers in these journals are published under an exclusive licence to publish, and authors are encouraged to follow our self-archiving policy by depositing their final version of the paper in an open access repository 6 months after publication. This conforms to all funder open access mandates worldwide*. See http://www.nature.com/authors/open_access/about_open_access.html.

[22] Under Horizon 2020, each beneficiary must ensure open access to all peer-reviewed scientific publications relating to its results.

[23] The term 'machine-readable electronic copy' means that the publications must be in a format that can be used and understood by a computer. They must be stored using text file formats which are either standardised or otherwise publicly known so that anyone can develop new tools for working with these documents.

[24] A repository for scientific publications is an online archive. The Open Access Infrastructure for Research in Europe (OpenAIRE) helps to determine what repository to choose (http://www.openaire.eu).

[25] Bibliographic data mining is more efficient than mining of full text versions.

[26] For example the Budapest declaration clearly states that '*The only constraint on reproduction and distribution, and the only role for copyright in this domain, should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited*. In: Suber P (2012) *Opening Access to Research*, August 24, 2012 published online http://www.berfrois.com/2012/08/peter-suber-opening-access-to-research/.

As Peter Suber states: *'In short, OA literature is digital, online, free of charge and free of most copyright and licensing restrictions. Ideally, the only restriction on use is an obligation to attribute the work to the author'.*[27]

There are several ways in which Open Access can be implemented:

- Gold Open Access refers to research being made available for free in its full, original form in the journal where it was published.[28] Gold open access journals can either be entirely open access, or they can be 'hybrid', in which subscription publications carry a subset of articles that are free for all to read.[29]
- Green Open Access refers to when Open Access is delivered by institutional or subject repositories or archives.[30] These hold the metadata about and copies of affiliated authors' works. In instances where the publisher policy allows it, this work is then made publicly available.[31]

To help provide Open Access to scientific information, OpenScienceLink has developed specific platform services for researchers, universities and publishers to help with Open Access publishing, peer review management, research and trend searches and evaluations. These will be discussed in more detail in the next chapter.

[27] Quoted from Peter Suber, Director of the Harvard Open Access Project, in Suber P (2012) *Opening Access to Research* retrieved from http://www.berfrois.com/2012/08/peter-suber-opening-access-to-research/.

[28] Open Access publishing means that an article is immediately provided in open access mode as published. In this model, the payment of publication costs is shifted away from readers paying via subscriptions.

[29] Martin Paul Eve (2014) Open Access and the Humanities, Cambridge Books Online, p. 8.

[30] 'Green' Open Access (also referred to as Self-archiving) means that the published article or the final peer-reviewed manuscript is archived (deposited) by the author - or a representative - in an online repository before, alongside or after its publication. In this case, they must ensure open access to the publication within a maximum of six months (twelve months for publications in the social sciences and humanities).Repository software usually allows authors to delay access to the article ('embargo period').

[31] A number of publishers allow authors to do this for journal articles. SHERPA/RoMEO is a useful tool for researchers to check publisher policies. Martin Paul Eve (2014)

# 3 OpenScienceLink platform services

## 3.1 Overview

Before analysing the relevant legal requirements for the OpenScienceLink platform, we will first provide a description of the services that will be developed within the platform. These services have been divided into separate pilots. For a detailed description of the technical aspects and developments within the OpenScienceLink platform services we refer to D.4.1.[32]

Hereafter we have summarized the acts that are relevant for the analysis of the legal requirements based on data protection and intellectual property law.

## 3.2 Open Access Data Journals Development

OpenScienceLink will provide the means for developing, publishing and managing data journals including a multitude of openly accessible data about experiments, trials and validation processes.

The availability of data journals will facilitate researchers in accessing scientific datasets in the scope of their research, while at the same time supporting them in relevant benchmarking processes (on the basis of specific datasets). A key characteristic of the data journals development process will be its ability to leverage semantic technologies/capabilities enabling the instant identification of the most appropriate datasets for researcher wishing to use or augment existing data journals.

### 3.2.1 Platform user registration

User registration is not obligatory but only necessary to be able to make use of some of the platform functionalities. To sign up for the platform the following personal information the following information is required: name, email and professional affiliation.

### 3.2.2 Research Data submission

In response to a call, researchers can submit their research paper and/or dataset for publication after peer- review. The paper needs to be in PDF format, but the datasets can be any format as long as this is described analytically, with detailed instructions and metadata on how the submitted dataset should be read, interpreted and re-used in future works.[33]

The uploaded data must comply with the legal and ethical requirements and may not infringe on intellectual property rights, hold illegal material or personal data.[34] Datasets must be anonymized before they are uploaded to the OpenScienceLink platform. Once the dataset is uploaded it becomes indexed

[32] D4.1 OpenScienceLink Platform Architecture and Background Platform Specifications, OpenScienceLink Consortium (2014)

[33] The metadata includes the author(s), an overall description of the data, purpose, conditions, features or parameters measured, references to related works that used this dataset, how the data could be used, and other notes or remarks that are important for re-use of the dataset.

[34] In case of biomedical research and clinical datasets, all variables that could reveal the identity of human subjects should be deleted (e.g. name, phone number, date of birth, insurance number etc.) so that datasets only include anonymous data.

and searchable for all Platform users, but not downloadable, which requires that the dataset is first reviewed and then published by the data journal.

### 3.2.3 Peer review process

After the datasets have been submitted, they must be evaluated for inclusion in the journal issue. For this purpose, the dataset is primarily connected using semantic enabled technologies with related literature, topics, and authors of the life sciences domains.[35] The publisher can then identify and assign certain datasets to certain reviewers for peer-reviewing. The reviewers evaluate the datasets according to different criteria, aided by the tools provided by the Platform.[36] When the review is finished, the Platform publishes the reviewers' comments, which can be viewed and discussed upon by other Platform users.[37]

### 3.2.4 Publishing

Once the review of the dataset is over, and the dataset is accepted for publication, it enters into a special status within the dataset pool, flagged as 'ready to be published'.[38]

The publisher uses the Platform in order to publish the selected datasets. The datasets and corresponding metadata can then be viewed and downloaded by the issue readers. Once a dataset is published, it is indexed with appropriate issue/volume numbers and pages, and turns into the 'published' stage. From this point on, the dataset is available to all of the Platform users for downloading and re-using in their research.

### 3.2.5 Dataset and Journal assessment and evaluation

The Platform provides the publisher with certain metrics for the assessment and evaluation of each dataset and the journal issue. The Platform also keeps track of papers that are referring to or have used each dataset. The main lever of semantically connecting the datasets with the rest Platform data entities (papers, authors, topics) is the list of the metadata and their values provided by the researchers during the uploading of the dataset. With this linking at hand, the publisher will be able to see and review statistics such as: number of times dataset appeared as result in queries, number of times dataset was

[35] Examples are text and datamining tools such as text annotation with ontology concepts.

[36] The reviewer may use the Platform's Semantic Filtering mechanism, which enables the semantic filtering of the initial semantic search results he is presented with. This filtering takes place through the ontologies incorporated in the OpenScienceLink Platform. The Platform allows the reviewer to apply a variety of ranking factors on the retrieved results, such as relevance, timeliness, quality of publications' fora, "authority" of author, etc., according to his/her preference.

[37] Depending on the publisher's permissions, other users of the Platform are now able to access the reviewers' comments, as well as the corresponding paper or dataset, and discuss them, by making new comments, depending on the permissions set by the publisher or editor. The comments submitted during the open-identity post-review discussion are annotated with terms from the ontologies incorporated into the OpenScienceLink platform. This way, the editor is able to easily navigate through the comments and filter them based on the ontology terms, and have a more efficient and fast way of having a clear view on the scientific community's feedback on the submitted reviews.

[38] This enables the publisher to select it as part of any future volume or issue for publication in the data journal.

downloaded, trendy topics that the dataset is related with, and other metrics, that can help the shaping of the dataset's evaluation, and significance in any life science field.

## 3.3 Data mining for Biomedical and Clinical Research Trends Detection and Analysis

The OpenScienceLink Platform through advanced data mining capabilities can help to discover and analyze research trends. This is done on the basis of data for the biomedical and clinical research field.

### 3.3.1 Trend Mining

All users that log into the Platform are able to view current research trends in the biomedical domain, including topical research areas and areas with the highest growth. The users can see per community, research field and subfields thereof, the measured trendiness over a pre-specified timeslot. The users are also able to view in a ranked list and respective graphs the highest trends, e.g., the scientific subfields that have accumulated the largest mass of scientific interest based on publications of papers and datasets in the field. Alternatively, the user enters a specific topic, requests for and is presented with its trend analysis over the pre-specified timeslot.

OpenScienceLink platform will provide automated generation of this information represented on a map showing the number of studies performed per city for a given scientific area.

### 3.3.2 Semantic searches

The users can perform, using the OpenScienceLink search tool, semantic searches across a variety of publicly available open access data sources, in order to identify relevant literature, research and research data.

## 3.4 Data mining for scientific collaborations

The platform provides for networking and collaboration of researchers and scholars working on similar fields and sharing similar interests. It facilitates the creation and tracking of networks of researchers, in a way similar to social networks. However, what makes the service different from conventional research information systems is the ability to infer relationships between researchers and research groups, including (in several cases) non-obvious, non-declared relationships. The platform service is therefore expected to open new horizons in the collaboration of researchers, scholars, research groups and research organizations. It will be primarily empowered by the advanced social networking and data mining capabilities of the OpenScienceLink Platform.

#### 3.4.1.1 Automated identification of groups of relevant researchers and relationships

Intelligent matching among researchers can be done, based on their dynamically generated research interests, as they are inferred through their published scientific work (including data sets, papers, articles, etc.). Matchings are filtered by taking into consideration their past and existing collaborations as well as their deduced acquaintance in social networks.

Similar matchings are made between researchers and research groups as well as communities based on their scientific topics in order to propose their collaboration and/or participation in them. For this purpose, the profile of the research group/community needs to be publicly available and accessible

through the Internet or, alternatively, the members of the group/community along with their scientific work.

## 3.5 Enhanced Research Evaluation Services

The platform service offers a new objective metrics of research and scientific performance, beyond conventional metrics, associated with conventional indices and impact factors. The service enables research sponsors, funding authorities and governmental agencies to shape their research strategies, researchers to be evaluated based on a multitude of factors representative of their productivity, impact and domain rather than through simplified means such as the number of citations within a time period, important research work, in terms of potential, to be brought forward, among others.

- Research evaluation

The Service uses mechanisms and Services that exploit the full range of openly accessible repositories of scientific work, the different metadata from the research data sources and the social networks in order to provide representative and complete evaluation of research outputs not only published scientific papers, but also data sets, researchers, research groups, communities, academic and/or research institutions and countries.

- The researcher's evaluation is based on the combination of the overall evaluation of all of their known research works.
- Journals are handled as an assortment of research works. The Platform calculates the overall evaluation of each of the research works that have been published in the journal individually, and combines them in order to provide the journal evaluation.
- The evaluation of a country is the result of the combination of the evaluation of each of the researchers in that country, and is calculated per research area (domain). The overall evaluation of a country is calculated based on the individual domains evaluation.

## 3.6 Summary OpenScienceLink platform services

To development of the OpenScienceLink platform and future implementation depends on the legal requirements that OpenScienceLink platform needs to comply with. In the next chapters we will discuss the relevant legal framework and corresponding requirements for the OpenScienceLink platform developments.

# 4 Privacy and data protection

One of the legal barriers for the OpenScienceLink project on open access to scientific information can be found in the legal framework for privacy and the protection of personal data. Based on the OpenScienceLink platform services, personal information may be subject to processing in the following situations:

(1) processing of personal information for user accounts on the OpenScienceLink platform

(2) processing of research data which may hold personal information (user uploads)

(3) further processing of personal information (re-use)

In the previous chapter we have presented an overview of the OpenScienceLink platform services. For a detailed description of the services we refer to D2.2. In the following sections we provide an overview of the European data protection directive and the requirements relevant for the OpenScienceLink platform.[39]

## 4.1 Overview of the Legal framework

The most important privacy and data protection regulations in the European Union are laid down in the following provisions:[40]

**Article 8 European Convention of Human Rights**

The European Convention for the Protection of Human Rights and Fundamental Freedoms (hereinafter 'Convention') has been ratified by all Member States of the EU. Under Article 8 (1), the Convention recognizes the fundamental right to privacy:[41] *Everyone has the right to respect for his private and family life, his home, and his correspondence.* [42]

[39] The more detailed analysis of the requirements as they apply to the specific platform services will be subject to a separate study under WP8: open access to research data.

[40] The object of the national laws on the processing of personal data is to protect fundamental rights and freedoms, notably the right to privacy, which is recognised both in Article 8 of the European Convention for the Protection of Human Rights and Fundamental Freedoms [signed in Rome on 4 November 1950 ('the ECHR')] and in the general principles of Community law.

[41] Lloyd, I (2014) Information Technology Law, Oxford University Press, 7th Edition.

[42] Recital 10 in the preamble to Directive 95/46 read as follows: the object of the national laws on the processing of personal data is to protect fundamental rights and freedoms, notably the right to privacy, which is recognised both in Article 8 of the European Convention for the Protection of Human Rights and Fundamental Freedoms [signed in Rome on 4 November 1950 ('the ECHR')] European Convention of Human Rights 2010. Retrieved from http://www.echr.coe.int/Documents/Convention_ENG.pdf

**Article 7 and 8 of the Charter of Fundamental Rights of the European Union**

'The respect for private and family life' contained in Article 7 and 'The right to data protection' in Article 8, are recognised as fundamental rights in the Charter of Fundamental Rights of the European Union (herein 'EU Charter').[43]

While Article 7 of the EU Charter is identical to article 8 of the Convention, Article 8 of the EU Charter explicitly establishes the right to data protection. It reads as follows:

*1. Everyone has the right to the protection of personal data concerning him or her.*

*2. Such data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law. Everyone has the right of access to data which has been collected concerning him or her, and the right to have it rectified.*

The EU Charter has binding legal status at EU level. As a result, the right to data protection must be balanced against other values and interests before the EU legislators and the European Court of Justice (hereinafter 'CJEU').[44]

## 4.2 Applicable EU data protection framework

The EU framework on data protection is composed of two main instruments: Directive 95/46/EC on the protection of personal data (Data Protection Directive), and Directive 2002/58/EC on the protection of privacy in the electronic communications sector (E-Privacy directive).[45]

### 4.2.1 E-Privacy Directive

The E-Privacy directive applies to "Electronic communications services" meaning a service normally provided for remuneration which consists wholly or mainly in the conveyance of signals on electronic communications networks, including telecommunications services and transmission services in networks used for broadcasting.[46] It excludes from its scope services providing or exercising, editorial control over content transmitted using electronic communications networks and services; information society services, as defined in Article 1 of Directive 98/34/EC, which do not consist wholly or mainly

[43] Charter of the Fundamental Rights of the European Union 2000, OJ C364/01. Retrieved from http://www.europarl.europa.eu/charter/pdf/text_en.pdf .

[44] Data Protection in the European Union: the role of National Data Protection Authorities (Strengthening the fundamental rights architecture in the EU II), FRA – EU Agency for Fundamental Rights, retrieved from http://fra.europa.eu/sites/default/files/fra_uploads/815-Data-protection_en.pdf, p. 18.

[45] The e-Privacy Directive applies to personal data processed in publicly-available electronic communications services in public communication networks (Article 3). Article 2(c) of the Framework Directive 2002/21/EC defines, 'electronic communications service' as a service provided for remuneration which consists wholly or mainly in the conveyance of signals on electronic communications networks, including telecommunications services and transmission services in networks used for broadcasting (e.g. ISPs, mobile and telephone operators, etc.).

[46] According to Article 2 sub c of the Framework Directive (2002/21/EC), which contains general definitions for the regulatory framework. As referred to in the Article 29 Working Party opinion: the Framework Directive explicitly excludes services providing or exercising editorial control over content.

in the conveyance of signals on electronic communications networks.'[47] We do note that the ePrivacy Directive includes some general provisions which are applicable not only to the electronic communication services but also to any other services. For example when OpenScienceLink would incorporate the use of so called cookie techniques it would have to comply with the relevant national regulations. [48] But, since these aspects are not characteristic for the OpenScienceLink platform, the E-Privacy Directive is not further discussed here.

### 4.2.2 Data Protection Directive

The objectives of the 95/46/EC Data Protection Directive are: to protect the fundamental right to data protection and to guarantee the free flow of personal data between Member States. [49]

Due to the technological progress and globalisation and because the Member States have implemented the Directive differently which resulted in divergence in enforcement, the directive is considered to no longer be adequate.'[50] On 25 January 2012 the European Commission, adopted a package for reforming the European data protection framework.[51] The reform is considered necessary '*to strengthen online privacy rights and boost Europe's digital economy*'.[52]

### 4.2.3 Data protection reform

Although the objectives and principles of the Data Protection Directive, which we will discuss hereinafter, remain valid, important changes are proposed.[53] The reform package includes two legislative proposals which will form the new legal framework:

---

[47] The E-Privacy directive is not subject to this documents analysis. The OpenScienceLink search services in the strict sense do not in general fall under the scope of the new regulatory framework for electronic communications of which the E-Privacy Directive is part.

[48] Article 5 (3) will then be relevant and OpenScienceLink needs to comply with the subsequent national regulations. The working part 29 in its Opinion 1/2008 on data protection issues related to search engines states that Article 5(3) of the E-Privacy Directive, to be read in conjunction with Recital 25 of the E-Privacy Directive, addresses the storage of information on the terminal equipment of users. Article 5(3) and Recital 25 of the E-Privacy Directive clearly stipulate that the storage of such information on the terminal equipment of users, i.e. cookies and similar devices, must be in accordance with the provisions of the Data Protection Directive.

[49] Directive 95/46/EC of the European Parliament and of the Council of 24.10.1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data (Data Protection Directive), Retrieved from http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:HTML

[50] The Commission report 'Evaluation of the implementation of the Data Protection Directive 'underlines that the implementation of the provisions of the Directive in national law has sometimes been unsatisfactory

[51] See furthermore The Commission Working Paper for the impact assessment which includes the study on the 'Evaluation of the implementation of the Data Protection Directive'(SEC (2012)72 final).

[52]European Commission, Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data, Brussels, 25 January 2012, COM (2012), 11 final. Retrieved from http://ec.europa.eu/justice/data-protection/document/review2012/com_2012_11_en.pdf.

[53] See for a report on status of the proposed Regulation, focusing in particular on the com- promise amendments adopted by the European Parliament in October 2013, and on the progress made in the Council of the EU.: C Burton, C Kuner, and A Pateraki, '*The Proposed EU Data Protection Regulation One Year Later: The Albrecht*

- (1) a proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation).

  The 'General Data Protection Regulation' is expected to be come into force by the end of this year.

- (2) a proposal for a Directive of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data by competent authorities for the purposes of prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and the free movement of such data. (Enforcement Directive)

  The status of the Enforcement Directive is not as far developed yet and is unlikely to enter into force before the end of the year.[54]

In the following sections our focus is on the data protection directive as it applies today and the proposed reform for as standing at the time of writing.

## 4.3 Scope of the Data Protection Directive

Until the newly proposed General Data Protection Regulation is finalised and enters into force, the Data Protection Directive is the core legal instrument governing the processing of personal data in the European Union.[55]

The application of the Data Protection Directive is defined under Article 3(1). The Data Protection Directive covers *the processing of any information relating to an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified , directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity wholly or partly by automatic means, and to the processing otherwise than by automatic means of personal data which form part of a filing system or are intended to form part of a filing system.*

Article 3(2) exempts two areas in which the Data Protection Directive is not deemed applicable:

- (1) Data processing carried out as part of activities falling beyond EU law (processing operations concerning public security, defense, national security, and activities in the area of Criminal Law).
- (2) Data processing carried out by a natural person in the course of a purely personal or household activity.

*Report'* Bloomberg BNA Privacy and Security Law Report (21 January 2013) and the follow up by the same authors aptly named 'The Proposed EU Data Protection Regulation Two Years Later.

[54] Given its uncertain position this report will not take the proposed directive into account, however we will monitor the developments and update when necessary.

[55] See http://ec.europa.eu/justice/dataprotection.

### 4.3.1 Personal data

Naturally, the Data Protection Directive only applies when 'personal data' are being processed.[56]

Article 2 (a) defines "personal data":

*'Any information relating to an identified or identifiable natural person'* where an 'identifiable' person is '*one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to this physical, physiological, mental, economic, cultural or social identity.'*[57]

In its Opinion on the concept of personal data, the article 29 Working Party (hereinafter Working Party) [58] has clarified the definition of personal data.[59] For the assessment of the concept of personal data four elements are relevant:

(1) 'Any information'
(2) 'Relating to'
(3) 'An identified or identifiable'
(4) 'Natural person'.[60]

As mentioned in the previous chapter the OpenScienceLink platform (1) allows its users to register for a personal account with their name and contact details, and (2) can be used for research(er) evaluation and trends analysis based on various sources and information from social media sites and publicly available websites and databases. In the following sections we will apply the criteria to understand the scope of 'persona data' and the implications for OpenScienceLink.

#### 4.3.1.1 Any information

The Working Party further defines the concept of what is *any information* by referring to the *nature* and *content* of the information.

- With respect to the *nature* of information the Working Party states that any sorts of statements about a person, true or false and objective or subjective can be considered relevant.

---

[56] Recital (23) of the Data Protection Directive: 'The principles of protection should apply to any information concerning an identified or identifiable person. To determine whether a person is identifiable, account should be taken of all the means likely reasonably to be used either by the controller or by any other person to identify the individual. The principles of data protection should not apply to data rendered anonymous in such a way that the data subject is no longer identifiable. According to the Article 29 Working Party 'When data dataset is truly anonymized and individuals are no longer identifiable European data protection law no longer applies'

[57] According to article 2 sub a) of the Data Protection Directive

[58] Article 29 of Directive 95/46/EC established a Working Party on the protection of individuals with regard to the processing of personal data, which became to be known as the Article 29 Working Party. This independent advisory board is comprised of representatives of the Member States Data Protection Authorities and issues interpretative documents without legal binding effects.

[59] Article 29 Working Party (2007), Opinion 4/2007 on the concept of personal data. Retrieved from http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2007/wp136_en.pdf

[60] Article 29 Working Party (2007).

- With respect to the *content* this means that information is not limited to information regarding the strict sphere of the individual's private and family life but may also be about whatever types of activities someone does, working relations or social behaviour. It also includes mundane, trivial or publicly available information and information can be in any format text image or sound file and contained in any medium. An email can include personal information for example. *Covering automatic processing of personal data within its scope it also includes data held in a computer memory.* [61]

#### 4.3.1.2 Relating to a natural person

The information needs *to relate to* an individual. The Article 29 Working Party states that *"data relates to an individual if it is data about a person"*. In the discussion on RFID the Working Party further states that *"[I]t refers to the identity, characteristics or behaviour of an individual or if such information is used to determine or influence the way in which that person is treated or evaluated"*. [62]

Sometimes there is no evident relationship between the data and the individual, but the data can still be considered personal data.[63] It is not necessary that the data "focuses" on someone in order to consider that it relates to him. The data can for example relate to another individual or even to an object and only indirectly to the individual.[64] In order to consider that the data to be considered as "relating" to an individual, one of the three following elements should be present:

1. "Content": information is given about a particular person, regardless of any purpose on the side of the data controller or of a third party, or the impact of that information on the data subject. For example if the results of a medical test is includes in a patient's medical file then this content links directly to the patient.

2. "Purpose": the data are used to evaluate, treat in a certain way or influence the status or behaviour of an individual. For example when devices are being tracked to monitor the performance of employees. This purpose allows linkage to the individuals who have been using these devices.

3. "Result": in the absence of a purpose or content the use of data is likely to have an impact on a certain person's rights and interest, taking into account all the circumstances.[65] For example tracking the position of cars on the road can help to provide real time traffic information, but could also be used by law enforcement to know if someone is speeding.

---

[61] Article 29 Working Party gives various examples that may help to determine whether personal information qualifies as personal data. The case of DNA data illustrates how information about the human body allows unambiguous and unique identification of a person. Article 29 Working Party (2007)

[62] As stated by Article 29 Working Party 'In general terms it must be information about a person.'. Article 29 Working Party (2007) p. 9.

[63] Article 29 Working Party (2007) p. 9.

[64] *Ibid.*

[65] Article 29 Working Party (2007) p. 26.

#### 4.3.1.3 Identified or identifiable

Article 2 of The directive states that: 'a natural person can be identified, *directly or indirectly*, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity". Identification is normally achieved through the 'specific' factors or pieces of information which the Article 29 Working Party refers to as "identifiers".

*"Directly" or "indirectly" identifiable*

A name is the most common example how an individual can directly be identified. But not only names qualify, if for example through the linking of several pieces of information a person can be identified, this combination of data could be protected as well.[66] Indirect identification depends on the context and typically relates to the phenomenon of "unique combinations", whether small or large in size.

The Court in the Lindqvist case stated that "*[R]eferring, on an internet page, to various persons and identifying them by name or by other means, for instance by giving their telephone number or information regarding their working conditions and hobbies, constitutes the processing of personal data*".[67]

The conclusion that IP addresses are personal data has been under debate. However the Article 29 Working Party also considered IP-addresses as identifiers, since they can - through the Internet Service Provider - be traced back to a person.[68] Also the proposed Regulation in its recital 24 specifies hat cookies and IP addresses constitute personal data unless they do not relate to an identified or identifiable individual.[69]

#### 4.3.1.4 Reasonable means

Recital 26 of the Directive reads that "*[t]o determine whether a person is identifiable account should be taken of all the means likely reasonably to be used either by the controller or by any other person to identify the said person.*"[70]

It is not considered "reasonable" when there is only a "*mere hypothetical possibility*" of identification. To consider whether a person is 'identifiable' the Article 29 Working Party has provided some additional factors that should be taken into account when considering if a person is identifiable.

---

[66] Kuner, C. (2003). European Privacy Law and Online Business. Oxford University Press, (51), p. 322

[67] Judgment of the European Court of Justice C-101/2001of 06.11.2003 (Lindqvist), §27.

[68] It has stated that "Internet access providers and managers of local area networks can, using reasonable means, identify Internet users to whom they have attributed IP addresses as they normally systematically "log" in a file the date, time, duration and dynamic IP address given to the Internet user. See WP 37: Privacy on the Internet - An integrated EU Approach to On-line Data Protection.

[69] See amongst other "The *Proposed EU Data Protection Regulation One Year. Later*: The *Albrecht Report*", *Bloomberg BNA Privacy* and *Security Law*. and the following study by the same authors.

[70] Note on reform: This is also mentioned in Recital 23 that "To ascertain whether means are
Reasonably likely to be used to identify the individual, account should be taken of all objective factors,
Such as the costs of and the amount of time required for identification, taking into consideration both available technology at the time of the processing and technological development".

These factors include:

- The cost of conducting identification;
- The intended purpose of the processing of this information;
- The way the processing is structured;
- The advantage expected by the controller;
- The interests at stake for the individuals;
- The risks of organization dysfunctions (such as breaches of confidentiality duties) and technical failures;
- The state-of-the-art in technology at the time of the processing and the possibilities for development during the period for which the data will be processed.

With regard to the last element, the state-of-the-art of technology, it has to be taken into account that with technological developments new identification possibilities may rise in the future. When developing a system that processes personal data for longer periods of time, these developments must be anticipated so that the system is able to adapt to new developments as they happen. Appropriate measures need to be incorporated when necessary since the reasonability test is considered to be dynamic.

### 4.3.2 Sensitive data

Some personal data is considered to have a more severe impact on individuals when being misused then other 'normal' personal data.[71] This is the case when the data holds personal 'sensitive' information. Sensitive information requires extra protection and may only be processed for specific purposes and under special conditions.

Art. 8 (1) of the Directive states that the processing of personal data *'revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, and the processing of data concerning health or sex life'* is generally prohibited.

According to the Article 29 Working Party advice paper the term "*revealing*" means that "*not only data which by its nature contains sensitive information is covered by this provision, but also data from which sensitive information with regard to an individual can be concluded*"[72].

Under Art. 8 (2), sensitive data may however be processed when falling under one of the following exemptions:

a) the data subject has given his or her explicit consent to the processing of the data;
b) the processing is necessary for the purposes of carrying out the obligations of the controller in the field of employment law;
c) the processing is necessary to protect the vital interests of the data subject or of another person;
d) the processing is carried out in the course of legitimate activities by a non-profit-seeking body with a political, philosophical, religious or trade-union aim;

[71] The Article 29 Working Party explains that misuse of these data could have more severe consequences on the individual's fundamental rights, such as the right to privacy and non-discrimination. Misuse of sensitive data, such as health data or sexual orientation (e.g. if publicly revealed), may be irreversible and have long-term consequences for the individual as well as his social environment. Article 29 Working Party (2007)
[72] See Article 29 Working party: (2011) Advice paper on special categories of data ("sensitive data")

e) or the processing relates to data which are manifestly made public by the data subject or is necessary for the establishment, exercise or defence of legal claims.

For the development of the OpenScienceLink platform services the focus is on research in the biomedical sector. Biomedical data as it is defined under the Directive may fall under the definition of sensitive data when it is data concerning 'health'.

Under the proposed reform of the Data protection directive a specific definition for health data is introduced: "*Data concerning health*" is defined as "*any information which relates to the physical or mental health of an individual, or to the provision of health services to the individual*".

Applying that definition health data also includes:

- information derived from the testing or examination of a body part or bodily substance, including biological samples
- information about disease risk
- Information about the actual physiological or biomedical state of the data subject independent of its source. [73]

Next to 'health data' the proposal also defines 'genetic data' and 'biometric data' as sensitive data:

- art 4(10) 'genetic data' means "*all data, of whatever type, concerning the characteristics of an individual which are inherited or acquired during early prenatal development*";
- art 4(11) 'biometric data' means "*any data relating to the physical, physiological or behavioural characteristics of an individual which allow their unique identification, such as facial images, or dactyloscopic data*"

It is especially important to clearly distinct between 'normal' personal data and sensitive personal data when applying the legitimate grounds for data processing. These legitimate grounds will be analysed more closely in section 4.4.5.

### 4.3.3 OpenScienceLink User profiles and biomedical data processing

If we apply the above mentioned elements to the OpenScienceLink services we can conclude that the information users provide (their name and email) is personal data. This information relates directly to the person who provides the information.

Given the purpose of identification of researchers for evaluation and networking the data mining and analysis of information may also fall under the scope of the Directive. The data gathered from a variety of sources is matched together and linked to an individual researcher. Therefore the data for the researcher's profiles and evaluations must be considered to be personal data.

The research data made available on the OpenScienceLink platform can come from various sources. However, OpenScienceLink only allows anonymized data to be uploaded to the platform. The anonymisation process is the responsibility of the platform user who must confirm before uploading the data to the platform it does not include personal data. If the research data is anonymized, the Data Protection Directive is no longer applicable. This is specified in Recital 46 of Directive 95/46/EC: "*the*

[73] Recital 34 of the Data Protection Directive

*principles of protection shall not apply to data rendered anonymous in such a way that the data subject is no longer identifiable".*

## 4.4 Anonymisation, Pseudonymisation and re-identifiability

For the OpenScienceLink project, anonymity is a key concern when the platform services involve the making available of clinical research data. The initial OpenScienceLink platform design would have allowed for researchers to upload and re-use any type of research data relevant in the biomedical field. However after evaluation of the platform services and legal framework it was decided that at this stage the platform should only allow research data when it is properly anonymized.

### 4.4.1 Anonymisation

The principle of storage limitation embodied in Article 6 of the Data Protection Directive as it will be explained, introduced the idea of data anonymisation at EU level. Anonymisation is a technique applied to personal data in order to achieve irreversible de-identification. If successful the resulting anonymized data can be processed without the Data Protection Directive being applicable.[74]

According to recital 26 of the Data Protection Directive *"to anonymize any data, the data must be stripped of sufficient elements such that the data subject can no longer be identified"*.[75] The Data Protection Directive is considered to have set a very high standard with respect to the anonymisation of personal data. As research has shown it may be almost impossible to fully avoid re-identification. Even if identification through a single data source is no longer possible (direct identification),by combining two or more data sources (indirect identification) data can become once again personal data.

A more realistic approach to the level of anonymisation that can be achieved, is stated by Article 29 Working Party which defines anonymous data as "*any information relating to a natural person where the person cannot be identified whether by the data controller or by any other person, taking account of all the means likely reasonable to be used either by the controller or by any other person to identify that individual*". The Article 29 Working Party argued for a flexible approach of the data protection laws, depending on the soundness of the technical and organizational measures adopted by the controller[76] as well as the probability of accidental re-identification of the data.[77]

As discussed previously under 4.3.1.4 a list of elements should be taking into account when assessing re-identification techniques.

---

[74] Recital 46: "Whereas the protection of the rights and freedoms of data subjects with regard to the processing of personal data requires that appropriate technical and organizational measures be taken, both at the time of the design of the processing system and at the time of the processing itself, particularly in order to maintain security and thereby to prevent any unauthorized processing; whereas it is incumbent on the Member States to ensure that controllers comply with these measures; whereas these measures must ensure an appropriate level of security, taking into account the state of the art and the costs of their implementation in relation to the risks inherent in the processing and the nature of the data to be protected".

[75] More precisely, the data must be processed in such a way that it can no longer be used to identify a natural person by using "all the means likely reasonably to be used" by either the controller or a third party.

[76] Article 17 of the Data Protection Directive.

[77] Article 29 Working Party (2007), (18) 26

The UK Information Commissioner's Office explains the complexity of the situation as follows: "different types of anonymised data have different vulnerabilities and pose different levels of re-identification risk: *"At one end of the spectrum, pseudonymised or de-identified data may be very valuable to researchers because of its individual-level granularity and because pseudonymised records from different sources can be relatively easy to match. However, this also means that there is a relatively high re-identification risk. At the other end of the spectrum, aggregated data is relatively low-risk [...] ","In general, the more detailed, linkable and individual-level the anonymised data is, the stronger the argument for ensuring only limited access to it*".[78] This data may be relatively 'safe' because re-identification risk is rather low. However, this data may not have the level of detail needed to support the data linkage or individual-level analysis that some forms of research depend on. Given the very different types of anonymised data that can be derived from personal data, it is important for data controllers to consider their disclosure options carefully, i.e. does the data need to be published or would limited access be appropriate?[79]

#### 4.4.1.1 Anonymisation process

A case-by-case analysis is necessary to see if the identification of a person based on 'all the means likely reasonable' is indeed no longer possible.[80]

The responsibility of anonymisation of research data lies with the researcher. However it may not always be evident or possible to anonymize data, for example in the case of statistical information, where in some cases, despite the fact that the information may be presented as aggregated data, the original sample might not be sufficiently large and other pieces of information may enable the identification of individuals.[81] OpenScienceLink therefore may consider to open up the platform for datasets with personal information. In the next section we will look at the possibility of reducing data protection issues by pseudonymizing data.

### 4.4.2 Pseudonymisation

Due to the difficulties of achieving anonymisation and the need to preserve valuable information in datasets, pseudonymisation may be used as a way to reduce data protection issues.

[78] This might be the case where it is necessary to use individual, record-level anonymised data to track particular individuals' movement through the education, employment and criminal justice systems Idem ICO [2012]

[79] See ICO (2012). Anonymisation: managing data protection risk: code of practice, (36) 108. Retrieved from http://www.ico.org.uk/~/media/documents/library/Data_Protection/Practical_application/anonymisation_code.pdf

[80] Article 29 Working Party (2013). Personal data, pseudonymous data and anonymous data: an Article 29 Working Party Future of Privacy subgroup discussion paper. *ICO*; Data Protection Commissioner. Data Protection Guidelines on research in the Health sector, (9). Retrieved from http://www.dataprotection.ie/documents/guidance/Health_research.pdf.

[81] Ohm, P. (2009). Broken promises of privacy: responding to surprising failure of anonymisation. *UCLA Raw Review 57*, (1715) 1703- 1777 and Article 29 Working Party (2013).

Pseudonymisation limits the linkability of a dataset with the original identity of a data subject by removing the direct identifier. However because the individual is still *indirectly* identifiable,[82] data protection regulations still apply.[83]

The effectiveness of the pseudonymisation procedure will depend on a great number of factors. In its opinion the Article 29 Working Party lists a number of them:

- at which stage it is used;
- how secure it is against reverse tracing;
- the size of the population in which the individual is concealed;
- the ability to link individual transactions or records to the same person.[84]

Other factors can be:

- the size of the data sets involved;
- the risks of an external hack;
- the likelihood that someone within the sender's organization would provide the key;
- the feasibility of indirect identification.[85]

It used to be generally acknowledged that Pseudonymisation presented a low but not insurmountable risk of re-identification.[86] However, due to our fast evolving global-technical environment, highly sophisticated data-matching software which is now more readily available, re-identification after Pseudonymisation has become much less difficult.[87]

Pseudonymisation has also been the topic of quite intense discussions in the context of the review of the Data Protection Directive. The main problem consists in the fact that there are different meanings to the concept of Pseudonymisation, making it very difficult to draw the line between anonymisation and Pseudonymisation. For instance, pseudonymous data can be data where a 'real' identifier (e.g. name and surname) is replaced by a 'false' identifier (e.g. hashed code number) or pseudonymous data could be an alternative form of personal identification, e.g. where online service companies use an IP address to target content at a particular device-user. Pseudonymous data can also be data that could potentially be combined with other data to produce personal data. This could however also apply to data from which all personal identifiers have been removed and where there is no reasonable likelihood of re-identification, i.e. anonymised data.[88]

---

[82] By linking different sets of data or by using corresponding lists for identifiers and their pseudonyms or two-way encryption algorithms. Article 29 Working Party (2013)

[83] Article 29 Working Party (2013b). Statement of the Article 29 Working Party on the current discussions regarding the data protection reform package, (1) 3. Retrieved at http://ec.europa.eu/justice/data-protection/article-29/documentation/other-document/files/2013/20130227_statement_dp_reform_package_en.pdf

[84] Article 29 Working Party (2007), (18) 26.

[85] Korff, D and Brown, I (2010), (48) 120.

[86] Nauwelaerts, W. (2007), Pseudonymisation of personal data in Pharmaceutical research: the art of disguising identities, *RAJ Pharma*, 595-597.

[87] Korff, D and Brown, I (2010), (50) 120.

[88] Article 29 Working Party (2013) Opinion 03/2013 on Purpose Limitation.

As a consequence, depending on the version of 'pseudonymous data', the data would be brought within or outside the scope of data protection law. [89]

### 4.4.3 Case study: key-coded data and the Clinical Trials Directive

One of the most common examples of pseudonymisation is to use key-coded data. Key-coded data constitutes data, which can only be linked to an identifiable person if you possess the decoding key. The data subject is only indirectly identifiable, via the key. Normally, the key – which makes the correspondence between the code and the common identifiers of the individual – is kept separately.[90]

A common situation in which key-coded data are used, are clinical trials with medicines. In this context, the legal framework set by the Clinical Trials Directive also applies.[91] In a clinical study, the medical professional or researcher ("investigator") testing the medicines will collect the information about clinical results on each patient, earmarking them with a code. Only this key-coded data will be reported to the pharmaceutical companies sponsoring the research. No personal data will be disclosed. The decryption keys will be held by the investigator to identify the patients for instance when the medicine turns out to pose dangers. The investigators are meanwhile bound by professional secrecy and good clinical practice to not reveal the identities of the data subjects.[92]

Does this case fall within the ambit of the Data Protection Directive? If the data could be considered to 'identify' a natural person –the data subject - this would entail that the Data Protection Directive applies. In practice, identification of the data subject is still possible via the key. Moreover, there is also the matter of 'accidental' matching of qualities of the data so that the data subject's identity is revealed.

According to our above analysis, purpose limitation is one of the key data protection principles contained in the Directive. In this case, the identification of the data subjects –in case treatment is necessary- is one of the main purposes of processing the key-coded data. For the pharmaceutical company sponsoring/ordering the clinical trial (the data controller) and all the parties involved in the possible identification, this key-coded data constitutes information relating to identifiable natural persons, which should be subject to the Data Protection legislation.[93] The Article 29 Working Party makes the distinction however as regards other data controllers who are processing the same set of data, but for who re-identification is completely and explicitly excluded. In the latter situation, the data protection legislation would not apply.[94]

---

[89] Kuner, C. (2007) p. 66. And for a different view Meier, P. (2010), *Protection des données*, Stämpfli p. 207.
[90] Article 29 Working Party (2007), (18) 26 and Korff, D and Brown, I (2010), (49) 120.
[91] Directive 2001/20/EC of the European Parliament and the Council of 4 April 2001 on the approximation of the laws, regulations and administrative provisions of the Member-States relating to the implementation of good clinical practice in the conduct of clinical trials on medicinal products for human use, *OJ* 121/34, 34. Retrieved at http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32001L0020:EN:HTML
[92] See also FAQ 14-7 of the Safe Harbor Scheme: Commission Decision of 26 July 2000 pursuant to Directive 95/46/EC of the European Parliament and of the Council on the adequacy of the protection provided by the safe harbour privacy principles and related frequently asked questions issued by the US Department of Commerce (notified under document number C (2000) 2441), *OJ* 215, 0007-0047.
[93] Article 29 Working Party (2007), p. 26.
[94] *Ibid.*

#### 4.4.3.1 Platform user related data: profile registration

The personal information that OpenScienceLink requires from its users (name, email) when they register for the platform are direct identifiers as they are a direct link to the platform user.

In the words of the Article 29 Working Party; *"computerised files registering personal data usually assign a unique identifier to the persons registered, in order to avoid confusion between two persons in the file. Also on the Web, web traffic surveillance tools make it easy to identify the behaviour of a machine and, behind the machine, that of its user. Thus, the individual's personality is pieced together in order to attribute certain decisions to him or her."* [95]

In its opinion on search engines the Article 29 Working Party concluded that also IP addresses and persistent cookies may contain unique identifiers and fall within the scope.[41] However this is a matter which has been discussed extensively and there is no consensus if IP addresses should indeed be qualified as personal data.[96] Still it is advisable to treat all IP information as personal data, to be on the safe side, as the Working Party advises unless it is possible to distinguish with absolute certainty that the data correspond to users that cannot be identified. [97]

It depends on the further developments whether this data will be collected and further processed under the OpenScienceLink services

#### 4.4.3.2 Datamining for trend and evaluation

OpenScienceLink collects information through text and datamining from publicly available websites and social networking sites. This data may refer directly to individual researchers. For example personal information about researchers taken from their professional profiles and publications. It can also be information about research institutes or objects such as articles and conferences which do not directly link to an individual. However personal data does not require the direct linking, matches such as the ones proposed by the OpenScienceLink service are based on linking of datasets. If these linked sets of data result in the identification of researchers it is likely they qualify as personal data.[98]

The Article 29 Working Party referred to the purpose element when discussing the reasonable means in relation to identification.[99] [W]here the purpose of the processing implies the identification of

---

[95] Article 29 Working Party Opinion 1/2008 on data protection issues related to search engines, adopted on 4 April, 2008, p. 14.

[96] For example EU Data Protection Supervisor (EDPS) Peter Hustinx affirms that all user activity, server logs and records of IP addresses could be classified as personal data. Details such as name, birth data, address, etc., do not need to be known from the data controller, as along as the processing of the IP address can be used to make him/her identifiable. ZDNet interview with Peter Hustinx, available at: http://news.zdnet.co.uk/security/0,1000000189,39540137,00.htm. And the Study of case law on the circumstances in which IP addresses are considered personal data. Which gives guidance on what circumstances should be taken into account when classifying IP addresses. SMART 2010/12 D3. Final report. p. 7.

[97] Study of case law on the circumstances in which IP addresses are considered personal data SMART 2010/12 D3. Final report. p. 24.

[98] However this is a matter that can only be determined in a case-by-case basis and needs to be interpreted in light of national legislation and case law.

[99] The Article 29 Working Party in its opinion, states that *[w]here the purpose of the processing implies the identification of individuals, it can be assumed that the controller or any other person involved have or will have*

individuals, it can be assumed that the controller or any other person involved have or will have the means "likely reasonably to be used" to identify the data subject. [100] If the purpose is to analyse trends in research the information may not qualify as personal data but as the information is also used to evaluate researchers and their output the consequences are that individuals based on the data may be identifiable.

#### 4.4.3.3 Re-identification and research data

Anonymisation and Pseudonymisation will continue to pose serious concerns about the possibility for re-use of scientific information. Paul Ohm states that "*—some techniques are very difficult to reverse—but researchers have learned more than enough already for us to reject anonymisation as a privacy-providing panacea.*"[101]

Apart from excluding personal information from the OpenScienceLink platform other measures such as minimizing not only the collection but also the storage time because "*once data have been collected and are stored, they are almost impossible to eradicate or truly, permanently anonymise.*"[102]

OpenScienceLink platform users must be made aware of the risks and standards for anonymisation and take into account the interpretation used by Data Protection Authority in their national jurisdiction. Good practices and guidelines, may help to avoid risks of re-identification and ensure anonymisation of the data being made available.

## 4.5 Processing of Personal data: Directive 95/46/EC

Whether the Data Protection Directive is applicable further depends on whether personal data are 'being processed'.

According to the current Data Protection Directive, Article 3 (1) the Directive applies to "*the processing of personal data wholly or partly by automatic means, and to the processing otherwise than by automatic means of personal data which form part of a filing system or are intended to form part of a filing system*".

Which includes: "*the collection, recording, organization, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, blocking, erasure or destruction"*.

The Data protection Directive does not apply to the processing of data:

- by a natural person in the course of purely personal or household activities;

---

*the means "likely reasonably to be used" to identify the data subject. In fact, to argue that individuals are not identifiable, where the purpose of the processing is precisely to identify them, would be a sheer contradiction in terms. Therefore, the information should be considered as relating to identifiable individuals and the processing should be subject to data protection rules.* Article 29 Working Party (2007) p. 16.

[100] One relevant factor, as mentioned before, for assessing "all the means likely reasonably to be used" to identify the persons will in fact be the purpose pursued by the data controller in the data processing. WP on personal data

[101] Ohm, P. (2009). p. 1716.

[102] Korff, D and Brown, I (2010), New Challenges to Data Protection - Final Report. European Commission DG Justice, Freedom and Security Report. Available at SSRN: http://ssrn.com/abstract=1636706 p. 48.

- In the course of an activity which falls outside the scope of Community law, such as operations concerning public security, defence or State security.[103]

The definition given of processing is very broad in scope and it is fair to say that almost all electronic handling of personal data, can be considered processing as meant by the Data Protection Directive. More specific the OpenScienceLink platform services of the storage and *otherwise making available* of the user profile data will qualify under the definition of processing. If personal data is taken from external sources for example from social media sites, and linked to the researcher's profile this will fall under the scope of processing.

Next thereto the OpenScienceLink project aims to help the researcher by providing access to research data which can be re-used for further research. This further use of data for research also constitutes as processing. However, currently the platform does not allow for the uploading and re-use of research data which holds personal information. Therefore we will not look at the question whether research data can be made available for re-use in detail. However, considering that the platform may develop services that allow the uploading of datasets including personal data, further analysis is proposed under work package 8 as to whether personal data processing for research purposes can be achieved under the data protection regulations.

## 4.6 Processing Actors

There are three important actors in the processing of personal data: data subjects, data controllers and data processors.

- **Data subject**

  The data subject is the person, the individual to whom the data relate. The data subject enjoys the protection of the EU data protection framework and is entitled to a series of rights.

  The data subject needs to be a natural person. The Data Protection Directive does not protect legal entities or organisations.

- **Data Controller**

  Article 2 (d) Data Protection Directive, defines the controller as "*the natural or legal person, public authority, agency or any other body which alone, or jointly with others, determines the purpose and means of the processing of personal data*".

- **Data Processor**

  Article 2 (e) Data Protection Directive, defines the processor as "*the natural or legal person, public authority, agency or any other body that processes personal data on behalf of the controller"*. According to the Article 29 Working Party to qualify as processor they need to be (1) a separate legal entity with respect to the controller and (2) processing personal data on the controller's behalf.

[103] Article 3. 2. Data Protection Directive.

Since a data processor is appointed by the data controller, the existence of data processors depend on a decision taken by the controller. The controller can decide either to process data within his organization or to delegate all or part of the processing activities to an external organization.[104] It has to be noted that in recent doctrine, one tends to argue that depending on the exact relation between the data controller and his delegate it may concern a controller – processor relationship or a co-controller relationship. When the decision on the means and purpose of the data processing is taken by one party, this party will be the data controller. The other party is considered to only act on behalf of the data controller. If however this decision on means and purpose is shared, co-controllership is accepted. The Article 29 Working Party also specified in its 2010 opinion on these concepts that "*the likelihood of multiple actors involved in processing personal data is naturally linked to the multiple kinds of activities that according to the Directive may amount to 'processing', which is at the end of the day the object of the 'joint control'*". [105] Other elements of distinction may include the degree of actual control exercised by a party, the image given to data subjects, and reasonable expectations of data subjects on the basis of this visibility.[106]

The qualification the processing parties as controller or processor is most important with respect to the fulfilment of the requirements under the directive and the delineation of liabilities.

- The data controller holds the primary duty to comply with the data protection requirements and bears the liability for violation of data protection.[107] Processors on the other hand can only be held responsible for non-compliance within his "share" of the processing activities and as long as the controller proves that he is not responsible for the event giving rise to the damage.[108] Exact delineation of liabilities should be specified in the controller – processor agreement, which is obligatory to draft according to Article 17, 3.
- Not surprisingly the controller is also the party the data subject should address when exercising his rights under Articles 10-12 and 14.[109] In second instance the data processor can be summoned by the data controller and the latter is obliged to reimburse the former.
- Article 4, 1 specifies that the place where the controller or his establishment is located is defining the applicable national law. In light of Article 4 of the Directive, processors and controllers must observe the standards of the applicable national data protection laws. The application of a Member State data protection law depends on: 1. the establishment of the data controller in the EU, 2. the use of equipment in a Member State by a non-EU data controller, and 3. Application of EU law based on public international law.
- The data controller needs to make a notification to the national Data Protection Authority[110]. The purpose of the notification is mainly to ensure transparency and provide information to the

[104] Article 29 Working Party Opinion 1/2010 on the concepts of "controller" and "processor" adopted on 16 February 2010, p. 1.
[105] Article 29 Working Party (2010), p. 18.
[106] Article 29 Working Party (2010), p. 12.
[107] Kuner, C. (2007), *op. cit.*, pp. 69-70.
[108] Article 23(2) Data protection Directive
[109] Article 29 Working Party (2010), p.5.
[110] Under Article 18 of the Data Protection Directive this is the responsibility of the controller of an automated filing system.

public.[111] The notified processing shall be kept in a public register of the supervisory authority. Any modifications to existing processing operations usually also need to be notified.[112]

Member States have the possibility, with regard to certain categories of processing operations, to provide for a simplification for such notification or for an exemption from notification.[113]

The OpenScienceLink platform allows its services to be used by a variety of different stakeholders.[114] If we look at the OpenScienceLink platform services we can conclude the following:

- The OpenScienceLink platform allows users to register with their personal information. This information is processed for the purpose and in the way OpenScienceLink has determined. In case the service is further developed by OpenScienceLink and OpenScienceLink acts as the controller responsible for the service of handling the use profiles.

- The OpenScienceLink platform service for evaluation and trend reports decides upon the tools it sues to text and data mine external sources including personal information from various sources. Some of the tools used are provided by third party actors. In this case OpenScienceLink is the controller who decides upon the data and what tools it wants to use and from whom. The actor providing these tools gets instructions on what data to collect but cannot decide upon the purpose of data collection.

- Hypothetically OpenScienceLink could be responsible as a processor for processing personal information within uploaded data sets.

  It can indeed be argued that the researchers are data controllers with regard to the sharing of the data through the platform. They decide on the data they collect and wish to share and for which purpose and they decide on the mean they prefer to use for this purpose. The researcher decides to:

  - Share his data sets on the OpenScienceLink platform
  - Share his data set for secondary processing of research data through the platform.

[111] Article 19 of the Data Protection Directive states that The notification must include at least: The name and address of the controller and of his representative; The purpose of the processing; Description of the categories of data subjects and of the data or categories of data relating to them; The recipients or categories of recipients to whom the data might be disclosed; Proposed transfers of data to third countries; General description allowing a preliminary assessment to be made of the appropriateness of the measures taken pursuant to Article 17 to ensure security of processing.

[112] In some cases it is possible that further investigation by the supervisory authority or by the data protection official is relevant and desirable. In that case, prior checking with the supervisory authority is mandatory. Article 20 of the Data Protection Directive.

[113] Chapter III Data protection officer; supervisory authority Section 36.

[114] Whenever a stakeholder contributes to the definition of the purpose and the means of the data collection, he shall be regarded as a controller. For an overview of all the different stakeholders involved and their role within the OpenScienceLink project we refer to OpenScienceLink Deliverable 2.1 .

In this case the researchers are required to comply with all the data protection requirements set forth by Directive 95/46, where applicable, and with national legislation. Given that the OpenScienceLink platform currently however only allows anonymous data sets, this last scenario is for the moment purely hypothetical. Nevertheless the terms of use of the Open Science Link platform clearly states that only the researcher can be held responsible for sufficiently anonymising data sets containing personal data.

*Note on current reform proposals:*

The proposed Regulation establishes a comprehensive responsibility and liability of the controller (art. 5 Regulation) on top of the legal requirements listed above. The controller must ensure and demonstrate for each processing operation the compliance with the provisions of the proposed Regulation.

Next, Article 19 of the proposed Regulation sets out that the Member States must ensure the compliance of the controller with the obligations arising from the principles of data protection by design and by default. It clarifies the position and obligations of processors, and adds that a processor that processes data beyond the controller's instructions is to be considered a co-controller. It also introduces the obligation for controllers and processors to maintain documentation of all processing systems and procedures under their responsibility.[115]

## 4.7 Complying with data protection principles and rules

The Data Protection Directive only allows the processing of personal data when it is done in accordance with the principles laid down in Articles 6, 7 and 8. Article 6 defines the core principles to lawful data processing. Article 7 and 8 define the legal grounds based on which respectively personal data and sensitive data may be processed. They regulate the so called legitimacy principle.

### 4.7.1 Article 6 Data Protection Directive: Principles relating to Data Quality

Article 6 states that personal data must be

- processed fairly and lawfully (lawfulness principle)

- collected for specified, explicit and legitimate purposes and not further processed in a way incompatible with those purposes (purpose limitation principle)

- adequate, relevant and not excessive in relation to the purposes for which they are collected and/or further processed (proportionality principle)

- kept in a form which allows identification of data subjects for no longer than is necessary for the purpose for which the data were collected or further processed (data minimization principle)

- accurate and kept-up-to-date (information quality principle)

[115] General Data Protection Regulation /* COM/2012/011 retrieved from at http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52012PC0011.

### 4.7.2 Fair and Lawful processing

#### 4.7.2.1 Fair

Fair processing requires data controllers to take into account the interests and reasonable expectations of data subjects.[116]

The limitation should, for example, prevent the use of individuals' personal data in a way (or for further purposes) that they might find unexpected, inappropriate or otherwise objectionable.[117]

- First, the collection and further processing of personal data should be carried out in a manner that does not intrude unreasonably upon data subjects' privacy nor interferes unreasonably with their autonomy and integrity. The concept of 'fairness' thus interlinks with the requirements of balance and *proportionality*.
- Second, a person should not be unduly pressured into supplying data on himself to a data controller or accepting that the latter uses the data for particular purposes. The data processing should be *transparent* for the data subject.[118]

Recital 38 of the Data Protection Directive clearly states that 'if the processing of data is fair, the data subject must be in position to learn of the existence of a processing operation and; where data are collected from him, must be given accurate and full information, bearing in mind the circumstances of the collection.'

#### 4.7.2.2 Lawful

The lawfulness of the processing requires: (1) data processing to be conducted only with legal basis and in compliance with all legal requirements, and (2) the right to data protection to be balanced against the interest of others in processing the data.[119]

Article 7 and 8 of the Data Protection Directive provide the legal grounds that legitimate respectively data processing of personal data and special categories of data.

### 4.7.3 Purpose limitation

Article 6(b) of the Data protection Directive describes the principle of purpose limitation: personal data may be collected only "*for specified, explicit and legitimate purposes and not further processed in a way incompatible with those purposes*".

According to the Article 29 Working Party, "*the principle of purpose limitation is designed to establish the boundaries within which personal data collected for a given purpose may be processed and may be put to further use*".[120] The Article 29 Working Party further specified that the principle has two components:

[116] Bygrave, L. (2002). Data Protection law: approaching its rationale, logic and limits. Kluwer Law International.
[117] Article 29 Working Party (2013), p. 70.
[118] Bygrave, L. (2002) p. 426.
[119] Kuner, C. (2007) p. 90.
[120] Article 29 Working Party (2013), p. 70.

- Purpose specification, meaning that personal data must be collected for 'specified, explicit and legitimate' purposes;
- Compatible re-use meaning data should not be further processed in a way incompatible with the original purposes.

#### 4.7.3.1 Purpose specification

The purpose must be specific, explicit and legitimate. Data subjects have the reasonable expectation that their data will only be used in accordance with the purposes for which the data was first collected. The very core of purpose limitation is to ensure this legitimate expectation is respected and that the information is not used in ways the data subject would not have agreed to or could not have anticipated. Also, it must encompass one of the legal grounds set forth under Article 7 of Directive 95/46/EC. A legitimate purpose embodies a specific justification for the data processing.[121] A general purpose such as 'scientific research' is not considered specific or explicit.

#### 4.7.3.2 Compatible use

Further processing of personal data for a different purpose for which it was collected, does not mean that it is incompatible. Even when data is collected for a certain purpose, another lawful form of processing can take place. The compatibility of re-use of personal data needs to be assessed on a case-by-case basis, taking into account all relevant circumstances.

In the specific opinion of the Article 29 Working Party devoted to the purpose limitation principle,[122] the following key factors determining the legitimacy of further processing were determined:

- The relationship between the purposes for which the personal data have been collected and the purposes of further processing;
- The context in which the personal data have been collected and the reasonable expectations of the data subjects as to their further use
- The nature of the personal data and the impact of the further processing on the data subjects;
- The safeguards adopted by the controller to ensure fair processing and to prevent any undue impact on the data subjects.

Processing of personal data in a way incompatible with the purposes specified at collection is against the law and therefore prohibited.[123] However, the Directive also specifies that – for as long as appropriate safeguards are adopted in national law – the further processing of data for historical, statistical or scientific purposes is not considered incompatible, even when the legitimate primary purpose was not for research purposes. [124]

[121] Lloyd, I (2014) *op. cit.*
[122] Article 29 Working Party (2013), p. 4.
[123] Article 6.b. Data Protection Directive.
[124] The example of further processing for research purposes was given by the Article 29 Working Party in their Opinion 06/2014 on the "Notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC"

### 4.7.4 Data Minimization

The data minimization principle is a fairly new principle, incorporating long standing ideas. It is as such not described in the current Data Protection Directive, but will be introduced by the proposals for reform. The idea is that by reducing the amount of personal data needed, data controllers can ensure better compliance with data protection requirements. Data minimization results in lesser interference in the private life of data subjects, enables greater balance of interests and proportionality, and reduces the risks of privacy and security breach. The proposed Regulation clarifies that the personal data may only be used "*if, and as long as, the purposes could not be fulfilled by processing information that does not involve personal data*".[125]

The principle of data minimization is twofold: it asks controllers to ensure the adequacy and relevance of the collected data while keeping the amount of personal data to the minimum necessary to achieve the purpose of the processing.[126] In that sense it would also enforce the storage limitation principle which states that personal data shall not be retained longer than necessary for achieving the purposes for which the data were first collected or further processed. The period of retention shall be specified before the start of the processing and is also part of the controller's duty of transparency, as embodied in the principle of fairness. This is because the duration of the data retention for periods beyond the required for the processing can violate the principle of proportionality: it places a burden on the data subject privacy without justification.[127] Since the Directive does not establish a fixed duration for legitimate data storage, this will be defined case by case. Clearly, the storage period shall depend on the purpose of retention and collection, on the nature more or less sensitive of the data, and on aspects of security, as data protection risks increase with time.[128] Consequently, applying the data minimization principle also implies that data controllers should erase or anonymise the personal data held under their control once they are no longer essential.

### 4.7.5 Information quality

Personal data should be valid with respect to what they are intended to describe, and relevant and complete with respect to the purposes for which they are intended.[129] Article 6(d), the principle of information quality stresses the obligation of data controllers to ensure the correctness of and reliability on the data they process. Data is regarded as being inaccurate when it is incorrect or misleading to any matter of fact.[130] Finally, it requires controllers to take all the appropriate measures to enable erasure or rectification of incorrect or incomplete data in light of the purpose for which they are collected and processed.[131] This duty is of permanent character and lasts for the duration of the processing, calling for data to be kept updated, when circumstances require so.

[125] Article 5.c of the General Data Protection Regulation
[126] Bygrave, L. (2002).p. 59.
[127] Meier, P. (2010), *op. cit.*, p. 272.
[128] Meier, P. (2010), *op. cit.*, p. 273.
[129] Bygrave, L. (2002). *op. cit.*, p. 62.
[130] Lloyd, I (2014) *op. cit.*, p. 117.
[131] Meier, P. (2010), *Protection des données*, Stämpfli p. 287.

## 4.8 Article 7 and 8: legitimacy principle

The processing of personal data is first of all restricted by the legitimacy principle. The legitimacy principle entails that for each act of data processing a legal ground needs to allow that specific act.

The processing of normal personal data is – as a rule – allowed, but only when one of the six legitimate grounds for the processing of normal personal data is applicable to that act of data processing. The processing of sensitive personal data is – again as a rule – not allowed, unless one of the five exemptions is applicable to the data processing act. Although in practice the distinction between the general permission to process personal data and the general prohibition to process sensitive data is very limited, it should be kept in mind that this distinction does imply a more restrictive view to the legal grounds under article 8.

### 4.8.1 Legitimate grounds for processing of normal personal data

Article 7 of the Directive contains an exhaustive list of six legitimate grounds for processing of personal information.

a) The data subject has unambiguously given its consent;
b) Processing is necessary of the performance of a contract to which the data subject is party;
c) for compliance with a legal obligation to which the controller is subject;
d) in order to protect the vital interests of the data subject;
e) for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller or in a third party to whom the data are disclosed;
f) for the purposes of the legitimate interests pursued by the controller or by the third party or parties to whom the data are disclosed, except where such rights are overridden by the interests for fundamental rights and freedoms of the data subject.

Member States cannot add new principles relating to the lawfulness of the processing of personal data to Article 7 or impose additional requirements that have the effect of amending the scope of one of the six principles provided for in that article.[132] In its judgement of 24 November 2011 in ASNEF and FECEMD, the CJEU held that Article 7(f) has direct effect limiting the freedom Member States have in implementing.[133]

### 4.8.2 Legitimate grounds for processing sensitive data

The processing of personal *sensitive* data is in principle prohibited unless

a. the data subject has given his *explicit consent* to the processing of those data, except where the laws of the Member State provide that the prohibition referred to in paragraph 1 may not be lifted by the data subject's giving his consent; or
b. processing is necessary for the purposes of carrying out the obligations and specific rights of the controller in the field of employment law insofar as it is authorized by national law providing for adequate safeguards; or

[132] This was determined in the joined cases C‑468/10 and C‑469/10, where the CJEU examined the legitimate character of further requirements to Article 7 of the Data Protection Directive.
[133] CJEU judgment of 24.11.2011 in cases C-468/10 and C-469/10 (*ASNEF and FECEMD*).

c. Processing is necessary to protect the vital interests of the data subject or of another person where the data subject is physically or legally incapable of giving his consent; or
d. processing is carried out in the course of its legitimate activities with appropriate guarantees by a foundation, association or any other non-profit-seeking body with a political, philosophical, religious or trade-union aim and on condition that the processing relates solely to the members of the body or to persons who have regular contact with it in connection with its purposes and that the data are not disclosed to a third party without the consent of the data subjects; or
e. The processing relates to data which are manifestly made public by the data subject or is necessary for the establishment, exercise or defence of legal claims.

The prohibition also shall not apply where processing of the data is required for the purposes of preventive medicine, medical diagnosis, the provision of care or treatment or the management of health-care services, and where those data are processed by a health professional subject under national law or rules established by national competent bodies to the obligation of professional secrecy or by another person also subject to an equivalent obligation of secrecy.

Subject to the provision of suitable safeguards, Member States may, for reasons of substantial public interest, lay down exemptions in addition to those laid down in paragraph 2 either by national law or by decision of the supervisory authority.

## 4.9 Applicable legal grounds for OpenScienceLink

Since OpenScienceLink does currently not allow the publication of personal data in the research data sets, we consider the prohibition to process sensitive data not relevant. Most relevant are however the legal grounds for processing the researcher's profile data. Two legal grounds are of particular importance for this use case:

- Article 7(a) the data subject has unambiguously given its consent;
- Article 7(f) processing is necessary for the purposes of the legitimate interests pursued by the controller or by the third party or parties to whom the data are disclosed, except where such rights are overridden by the interests for fundamental rights and freedoms of the data subject.

For the sake of completeness and in order to provide some insight for future deployment of the OpenScienceLink platform, we also briefly look into the use of personal data for research purposes.

### 4.9.1 Consent

The directive states that for consent to be valid it has to be freely given, specific and informed and unambiguous.[134] In its opinion, the Article 29 Working Party has adopted the following:

- Only statements or actions that indicate the data subject's agreement may constitute valid consent. Mere silence or inaction (opt-out) typically will not be viewed as valid consent, especially in an online context. For example, default privacy settings used by online social networks, default Internet browser settings or pre-ticked boxes do not qualify as valid consent.
- Consent must be given prior to the data processing, after providing notice to the data subject. The notice should be provided in clear and understandable language.

[134] Note that to rely on consent for the processing of sensitive data this consent has to be 'explicit 'consent.

- From an accountability perspective, data controllers should implement mechanisms to prove that they have obtained data subjects' valid consent.
- Reliance on consent does not relieve data controllers from their obligation to comply with other EU data protection requirements for lawful processing of personal data, such as the principle of proportionality.
- If a data subject withdraws his or her consent, the data controller must delete any personal data pertaining to the data subject unless there is another legal basis that justifies continued storage of the data. [135]

The new proposed regulation has clarified the requirements for consent. 'the data subject's consent' means any freely given specific, informed and explicit indication of his or her wishes by which the data subject, either by a statement or by a clear affirmative action, signifies agreement to personal data relating to them being processed'. The new Article 7 sets the following conditions to informed consent:

1. The controller shall bear the burden of proof for the data subject's consent to the processing of their personal data for specified purposes.
2. If the data subject's consent is to be given in the context of a written declaration which also concerns another matter, the requirement to give consent must be presented distinguishable in its appearance from this other matter.
3. The data subject shall have the right to withdraw his or her consent at any time. The withdrawal of consent shall not affect the lawfulness of processing based on consent before its withdrawal.
4. Consent shall not provide a legal basis for the processing, where there is a significant imbalance between the position of the data subject and the controller.

For the processing of personal data with regards to the OpenScienceLink platform user profiles, Consent will be relied on as the legitimate ground.

### 4.9.2 Legitimate interest

Article 7(f) is the last of six grounds for the lawful processing of personal data.

To be considered as legitimate, an interest must cumulatively fulfil the following conditions:

- be lawful (i.e. in accordance with EU and national law);
- be sufficiently clearly articulated to allow the balancing test to be carried out against the interests and fundamental rights of the data subject (i.e. sufficiently concrete);
- Represent a real and present interest (i.e. not be speculative).

Additionally, there may not be other less invasive means to reach the identified purpose of the processing and serve the legitimate interest of the data controller. In other words it must be necessary to achieve the interest pursued.

Finally the Article 29 Working Party set some balancing factors to take into consideration to determine whether Article 7(f) may be relied upon as a legal ground for processing are:

[135] Working Party 29 Opinion 15/2011 on the definition of consent, retrieved from http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2011/wp187_en.pdf

- the nature and source of the legitimate interest;
- whether the data processing is necessary for the exercise of a fundamental right, is otherwise in the public interest, or benefits from recognition in the community concerned;
- the impact on the data subject;
- their reasonable expectations about what will happen to their data;
- the nature of the data and how they are processed;
- additional safeguards which could limit undue impact on the data subject, such as data minimization, privacy-enhancing technologies;
- increased transparency, general and unconditional right to opt-out, and data portability. [136]

Examples on how to apply the assessment are also provided by the Article 29 Working Party's Opinion. Especially relevant for OpenScienceLink are the examples referring to the research using personal data and Reuse of publicly available data.[137]

### 4.9.3 Legitimate ground for the purpose of research

That processing for research purposes may be legitimate has been confirmed by the introduction of the legitimate ground in art 6, 1, b: "*further processing of data for historical, statistical or scientific purposes shall not be considered as incompatible provided that Member States provide appropriate safeguards*". Consequently the current situation however relies on the implementation of this provision by the Member States. Based on the above cited article 6 (2) and article 32 of the Data Protection Directive, Member States may, subject to suitable safeguards, allow that data kept for the sole purpose of historical, statistical and scientific research.138 Some Member States have indeed implemented provisions that allow further use of personal data for research purposes, but regulations are scattered and far from harmonized.

*Note on current reform proposals*

[136] Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC

[137] see example 7, 19 and 23: Rating of politicians by NGO where Even if the impact on politicians concerned may be significant, the fact that processing is based on public information and in relation to their public responsibilities makes, with a clear purpose of enhancing transparency and accountability, the balance tips in the interest of the controller, Opinion 06/2014

[138] For example the German data protection act reads as follows:
(1) Personal data collected or stored for scientific research purposes may be processed or used only for such purposes. (2) The communication of personal data to other than public bodies for scientific research purposes shall be admissible only if these undertake not to process or use the communicated data for other purposes and to comply with the provisions under 3 (3) The personal data shall be depersonalized as soon as the research purpose permits this. Until such time the characteristics enabling information concerning personal or material circumstances to be attributed to an identified or identifiable individual shall be stored separately. They may be combined with the information only to the extent required by the research purpose. (4) Bodies conducting scientific research may publish personal data only if a)the data subject has consented or this is indispensable for the presentation of research findings on contemporary events

Relevant for the further developments of the OpenScienceLink platform services is that the proposed reform regulation contains a specific section on data processing for research with *archiving, scientific, statistical or historical purposes.*[139] Under the proposed Regulation personal data may be used for these purposes if the research purposes cannot be obtained by anonymous data and, in that case, the data enabling the attribution of the information to a particular person is kept separately from the other information, to the extent possible. With respect to research publications however, the data subject needs to give consent to the publication of the research results. The publication of personal data should be necessary to present research findings, or to facilitate research, or the data subject has made the data public.

The rationale behind this exemption is that the collection of personal data has been done based on legitimate grounds so for the re-use for the above mentioned research purposes no other legitimate ground is required. In practice this would mean that OpenScienceLink users may re-use existing datasets that include personal information for the purpose of research under the conditions mentioned above. They would not have to obtain additional consent from the data subjects for their re-use purposes.

## 4.10 Data subjects' rights

As important as a core principle of data protection law is that data subjects should be able to participate in, and have a measure of influence over, the processing of data on them by other individuals or organisations.[140] OpenScienceLink must comply with the above mentioned data protection principles and respect the rights of the data subjects.[141]

### 4.10.1 Right to information

#### 4.10.1.1 Article 10

The controller or his representative must provide a data subject from whom data relating to himself are collected with at least the following information, except where he already has it:

(a) The identity of the controller and of his representative, if any;

(b) The purposes of the processing for which the data are intended;

(c) Any further information such as

- *the recipients or categories of recipients of the data,*
- *whether replies to the questions are obligatory or voluntary, as well as the possible consequences of failure to reply,*
- *the existence of the right of access to and the right to rectify the data concerning him*

[139] Article 83 of the General Data Protection Regulation.
[140] Bygrave, L. (2002), (63) p. 426
[141] Article 10-15 Data Protection Directive.

*In so far as such further information is necessary, having regard to the specific circumstances in which the data are collected, to guarantee fair processing in respect of the data subject.*

On the condition that Member States foresee appropriate safeguards, exceptions can be made to the duty to inform when the effort of informing the data subjects would be impossible or disproportionate.[142] In the current Directive, it is stressed that this could in particular be the case for the processing of personal data for statistical purposes or for the purposes of historical or scientific research.

*Note on current reform proposals*

The proposed Regulation stresses the transparency principle and introduces the obligation on controllers to provide transparent and easily accessible and understandable information.

### 4.10.2 Right of access, correction and deletion

The data subject has the right to challenge the accuracy of the data and to provide correct information. The access process should be timely, inexpensive, simple, providing a mechanism for verification of the data and a means by which corrections and objections may be recorded. The data subject can also ask the erasure or blocking of the data of which the processing does not comply with the provisions of the Directive, in particular because of the incomplete or inaccurate nature of the data.[143]

*Note on current reform proposals:*

The proposed Regulation introduces the data subject's right to data portability, i.e. to transfer data from one electronic processing system to and into another, without being prevented from doing so by the controller. As a precondition and in order to further improve access of individuals to their personal data, it provides the right to obtain from the controller those data in a structured and commonly used electronic format.[144]

#### 4.10.2.1 Right of integrity and security

OpenScienceLink must implement appropriate technical and organizational measures to ensure the confidentiality and security of the processing.[145] In particular this means taking appropriate measures to "*protect personal data against accidental or unlawful destruction or accidental loss, alteration, unauthorized disclosure or access […] and against all other unlawful forms of processing*".

The data subject has the right to know the extent to which the data will be secured. To ensure data integrity, OpenScienceLink must take reasonable steps, which includes only using reputable sources for

[142] Article 11 of the Data Protection Directive.
[143] Article 12 of the Data Protection Directive.
[144] For instance, if an individual who posted photos on a photo-sharing site decides that he wants to use another service provider, he should be able to transfer all these photos where technically possible, without any obstacles or loss. European Commission (2012). How will the EU's reform adapt data protection rules to new technological developments ? Retrieved from http://ec.europa.eu/justice/data-protection/document/review2012/factsheets/8_en.pdf
[145] Based on Article 16 and 17 of the Data Protection Directive.

its data collections and crosschecking data against multiple sources, providing individuals access to data, and destroying access data.

Security of the data would include both the management of and the technical measures to protect against loss, unauthorized access, use and disclosure of the data.[146] To understand better how to do this in practice several national data protection authorities and governmental agencies have issued guidelines as to the types of organizational and technical measures data controllers should consider when handling personal data.[147]

#### 4.10.2.2 Confidentiality and role-based rights

A first security objective following from the controller's security obligation is to maintain the confidentiality of information. It means that one should keep the content of information secret from all entities except those authorized to access it. There are numerous approaches to providing confidentiality, ranging from physical protection to the use of access control and cryptographic algorithms. In addition to safeguarding the confidentiality of information, the processing capabilities (read, write, modify,) of each entity should be limited to that which is necessary to realize the goals of the processing. This follows from a combined reading of the controller's security obligation and the proportionality principle. These requirements apply not only at the level of each organization, but also at the level of each individual user.[148]

#### 4.10.2.3 Integrity and authenticity

OpenScienceLink is required to integrate appropriate security policies to safeguard the *integrity* and *authenticity* of the data. The platform users must be able to establish whether the information comes from an authoritative and authorized source. Especially with respect to data from authoritative sources, the integrity and authenticity of the data flowing to and from these sources should be protected, e.g. through use of data origin authentication protocols (which also serve to establish their integrity during transmission). Relying parties should only process personal data further if there is sufficient certainty as to its origin and integrity (i.e. upon verification that it emanates from the intended source and has not been subject to manipulation).

With respect to the user profiles a researcher can only register by name and email, which will be used to verify their account and in case of loss of passwords.

#### 4.10.2.4 Accountability

Although the meaning of this concept varies across disciplines, as a security objective the term refers mainly to non-repudiation or general auditability requirements.[149] In this context, accountability denotes

---

[146] Kosta, E (2012). Search engines and data protection- Google under the microscope of European regulatory authorities, (8) 46. Retrieved from http://people.cs.kuleuven.be/~bettina.berendt/teaching/Privacy12/Kosta.pdf

[147] Kuner, C.(2007) p.288 et seq.

[148] Van Alsenoy B., Kindt, E. and Dumortier, J, (2010) Privacy and Data Protection Aspects of e-Government Identity Management p 258

[149] S. Pearson and A. Charlesworth, 'Accountability as a way forward for Privacy Protection in the cloud', in M.G. Jaatun, G. Zhao and C. Rong (Eds.), *Cloud computing. First International Conference. CloudCom 2009*, December 2009, Beijing, China, p. 134.

that a system or protocol has been designed in such a way that relevant events can be reconstructed (e.g. in order to assess policy compliance, to determine the cause of system failure) or that plausible deniability has been diminished. In other words, it refers to the need to enable traceability of the actions by entities in such a way that they might be held answerable ('called to account') if they engage in unauthorized processing activities.[150]

#### 4.10.2.5 Availability

Data controllers are also under the obligation to protect personal data against accidental destruction or loss. This requirement can be approximated to the security objective of *availability*, which can be described as the property of being accessible and usable upon demand by any authorized entity.

### 4.10.3 Right of enforcement and redress

The data subject has the right to seek legal relief to protect his privacy rights. This is listed in Article 22 and 23 of the Data Protection Directive.

*Note on current reform proposals*

The proposed Directive obliges controllers to provide procedures and mechanisms for facilitating the exercise of the data subject's rights. This includes the requirement that the exercise of the rights shall be in principle free of charge.

## 4.11 Summary Data Protection requirements

**Implementation of the Data Protection Requirements**

| *Data Protection compliance* |
|---|
| |
| *Allocation of roles and responsibilities* |
| With regard to the researcher's profiles OpenScienceLink determines the purpose and means of the processing of the personal data and can be qualified data controller.<br>Consequently, OpenScienceLink is primarily responsible and liable for the legality of the processing and the fulfillment of obligations towards the National Data Protection Authority (*NDPA*) and the data subjects.<br>In case third party services are used for processing personal data this relationship will be governed by a contract wherein the different obligations of the different parties are well indicated. |
| *Ensure legitimacy of processing* |

[150] See also European Court of Human Rights, I *vs. Finland*, 17 July 2008, par. 41-44

For the *OpenScienceLink platform*, the ground for processing will be the *data subject's* consent (e.g. via an electronic consent form).

*At the moment the platform only allows for processing of non-personal research data.*
*If however after review open access can be given to research data which holds personal data an access system with restricted access may be developed to allow use for research purposes.*

*Requirements for consent*

The consent must be:

- Unambiguous
- Specific, distinctive and intelligible
- Based on accurate, full and understandable information
- Genuine and freely given, absent of any pressure
- Given before any processing occurs
- Given with the option to withdraw his or her consent and stop any further processing of the personal data

These requirements have been taken into account when developing the profile registration form.

*Data Quality Principles*

The personal data and the processing must comply with the following data quality principles for processing personal data:

Personal data must be:

- Processed fairly and lawfully
- Collected only for specified, explicit and legitimate purposes and not further processed in a way incompatible with those purposes
- Adequate, relevant and not excessive in relation to the purposes for which it is collected and/or further processed
- Deleted or rendered anonymous when no longer necessary for the purpose
- Securely deleted, i.e. the data should not be retrievable after deletion
- Accurate and up to date.

*In its privacy notice and/or terms of use OpenScienceLink Platform states the purpose of the data processing, taking into consideration if the data being processed is both adequate and non-excessive to achieve the purposes of the OpenScienceLink platform. To make the processing of personal data lawful OpenScienceLink will rely on informed consent for the processing of user data from data subjects directly (when datasets with personal data are uploaded by users they have to state that this is in compliance with the data protection regulations. Confirmation of consent forms may become*

| |
|---|
| *mandatory before upload). Subsequent use will be limited to those articulated purposes. When asked for information it will clearly indicate whether information is optional or required.*<br>*OpenScienceLink will limit storage duration whenever possible while also ensure that the user data is readily available to authorized entities as long as it is necessary.*<br>*Procedures will be put in place to be able to delete information securely when no longer necessary for the purpose on how to report and deal with suspected inaccuracies. OpenScienceLink in its function as platform will rely on the data controller to delete or render the information anonymous when no longer necessary for the purpose and keep the information accurate and up to date.* |
| *Transparency and administrative requirements*<br>The data controller and the *OpenScienceLink infrastructure* must ensure the data subject's rights. |
| Being identified as data controller OpenScienceLink must<br>• Meet all notification and authorization requirements for data processing that may derive of national law of the competent Member State.<br>• Provide the data subject with sufficient information, at least<br>(1) the identity of the controller, (2) the categories of data to be processed, (3) whether the info is obligatory or voluntary, (4) the purpose of the processing, (5) the recipients of the data and (6) the further rights to access and to rectify.<br>• Provide the data subject with the right to obtain intelligible information without expense or excessive delay<br>• Support the data subject's rights to legitimately rectify, reply, revoke, erase or block his or her personal data<br>• Ensure that it has filed a notification to the National Data Protection Authority *prior* to each activity involving personal data processing. Be prepared for any requests by the Commission for additional items of information related to the processing.<br>• The consent of the data subject must be obtained *prior* to any processing of his personal data. A versioning and archiving system must be in place for the informed consents given by data subjects to enable later verification that appropriate notice was given.<br>*The OpenScienceLink platform in its current constellation falls under German law. Sufficient information about the processing of personal data, its purpose and the rights of the data subject to rectify is provided in the privacy notice and referenced on the OpenScienceLink website when relevant.* |
| *Security and confidentiality requirements* |
| The *OpenScienceLink infrastructure* shall implement advanced organizational and technical security measures to ensure confidentiality, integrity and authenticity. *This will be addressed during the further developments of the pilots and platform website* |

# 5 Intellectual property rights

## 5.1 Introduction

The OpenScienceLink platform services are developed to facilitate the publication, sharing, linking, review and evaluation of research results, based on open access to scientific information. The project focus is on research data in the biomedical field.[151]

The purpose of this chapter is to analyse the requirements based on the European legal framework for copyright and database protection.

*Intellectual property rights should allow the inventor or creator to derive a legitimate profit from his/her invention or creation, the widest possible dissemination of works, ideas and new know-how,* and *not hamper freedom of expression, the free movement of information, or the protection of personal data, including on the Internet.*[152]

The following scenarios provide an overview of the platform services relevant for this chapter's analysis of the legal framework. These scenarios cover most of the acts and works under the OpenScienceLink project. This Chapter is an update of D3.2 to reflect the changes in the developments of the OpenScienceLink project as well as the legal framework. We refer to D3.1 for a detailed description of the pilot services. [153]

*Scenario 1*

*Users can upload their research datasets and publications to the platform for peer review when there is an open call. This research data including the metadata will be made available on the platform under an open access license and can be downloaded for further use. In the first part of this chapter we will address the issue of copyright for research data. We will focus on the scope of copyright protection, i.e. the scope of the exclusive rights, limitations and exceptions as well as the issue ownership rights.*

*Scenario 2*

*Users can use the platform services to find research trends, data and to identify other researchers for collaborations. OpenScienceLink will use linking and data analysis such as text and data mining applied to external databases. In the second part of the chapter we will focus on the act of using third party content for the services of the platform. More specifically we will examine whether research databases*

[151] Clinical research refers to medical research combined with professional care, see more specifically the MRC Good Clinical Practice in Clinical Trials guidance. And WMA Declaration of Helsinki - Ethical Principles for Medical Research Involving Human Subjects.

[152] Corrigendum to Directive 2004/48/EC of the European Parliament and of the Council of 29 April 2004 on the enforcement of intellectual property rights.

[153] D3.1 Detailed Specification of OpenScienceLink Pilot Scenarios and Services, OpenScienceLink Consortium (2014).

*such as the PubMed repository are protected under the Database directive and what use requires permission from the respective database owner.*

In the following sections we will outline the legal framework within which OpenScienceLink can further develop its platform services. We will successively discuss (5.3) Copyright protection for original datasets, metadata and publications, and (5.6) Protection for databases, thereby distinguishing between (5.7) Copyright protection for original databases and (5.8) the *sui generis* right for databases. At the outset, we will provide a short overview of the applicable legal framework.

## 5.2 The Legal framework

The current legal framework is a combination of international treaties, European directives and national legislation.

### 5.2.1 International

The Berne convention for the Protection of Literary and Artistic Works is still considered to be the main international reference for copyright protection.[154] The Berne convention has set the minimum requirements of protection together with a set of basic principles. These minimum standards were improved by the TRIPs agreement[155] and, with a view to their application in the digital age, by the WIPO Copyright Treaty of 1996.[156]

The European Union has transposed the international obligations contained in these treaties into its legal framework.

### 5.2.2 European framework for Copyright protection

The European Union legal framework on copyright protection consists of the following directives: [157]

- the 1991 Computer Programs Directive (91/250/EEC, codified by 2009/24/ EC),
- the 1992 Rental and Lending Directive (codified by 2006/115/E),
- the 1993 Satellite and Cable Directive (93/83/EEC),
- the 1993 Term Directive (codified by 2006/116/EC, amended by 2011/77/EU),
- the 1996 Database Directive (96/9/EC) and
- The 2001 Resale Right Directive (2001/84/EC)
- The 2001 Directive on the harmonisation of certain aspects of copyright and related rights in the information society (InfoSoc Directive)[158]. The international obligations such as those laid

[154] Berne Convention for the Protection of Literary and Artistic Works (first signed in 1886, last amended in 1979) Another relevant treaty is the Rome Convention for the protection of Performers, producers of phonograms and broadcasting organisations (1961).

[155] Agreement on Trade Related Aspects of Intellectual Property Rights (TRIPS) (1994).

[156] See also the comparable WIPO Performances and phonograms Treaty (1996) in respect of the protection of neighbouring rights.

[157] All the relevant legislation forming the EU legal framework ("acquis") can be accessed online via http://ec.europa.eu/internal_market/copyright/acquis/index_en.htm

[158] Directive 2001/29 on the harmonisation of certain aspects of copyright and related rights in the information society, OJ 167/10. Retrieved at http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32001L0029:EN:HTML

down in the Berne convention and subsequent treaties have been transposed into Community law mainly by this InfoSoc Directive.

- The 2012 Orphan Works Directive (2012/28/EC)
- The 2014 Directive on collective management (2014/26/EC)

For the purpose of the OpenScienceLink project, the 2001 InfoSoc Directive clearly constitutes the main instrument with respect to the EU legal framework on copyright. It will therefore be the focus of our analysis.

In its preamble it states that

*[W]without harmonization at Community level, legislative activities at national level…, might result in significant differences in protection and thereby in restrictions on the free movement of services and products incorporating, or based on, intellectual property, leading to a refragmentation of the internal market and legislative inconsistency.*

Despite its many improvements to the EU copyright framework, its aim of harmonization has not been fully achieved.[159] As a consequence copyright reform has been proposed, the details of which will only be made public in the second half of 2015. We will hereafter briefly discuss some of the major points of the intended reform[160], although the latter has to be considered with reservation.

### 5.2.3 European copyright reform

The initiative to modernise the EU copyright framework was launched in May 2011 in the European Commission's strategy on "A Single Market for Intellectual Property Rights"[161] and pursuant to actions in the Commission's Digital Agenda for Europe[162]. In December 2012, the Commission then published a "Communication on Content in the Digital Single Market".[163]

The review must ' *achieve a modern framework that fosters innovative practices, creativity, cultural diversity, new business models, guarantees effective recognition and remuneration of rights holders, and enhances legal offers for end users while tackling piracy more effectively.'*

The review, including public consultations[164] and various studies, was completed in 2014. As was said before, concrete legislative reform proposals are expected to be published later this year. This may affect

[159] see Study on the application of Directive 2001/29/EC on copyright and related rights in the information society (http://ec.europa.eu/internal_market/copyright/docs/studies/131216_study_en.pdf)

[160] A public consultation was part of the program to review the current legal framework on copyright. Topics identified in the Consultation ranged widely from territoriality in the Single Market, harmonization, limitations and exceptions to copyright in the digital age, fragmentation of the EU copyright market to how to improve the effectiveness and efficiency of enforcement while underpinning its legitimacy in the wider context of copyright reform. Questions included everything from linking, browsing, copyright term duration, limitations and exceptions for cultural heritage institutions to 'user generated content'. http://ec.europa.eu/internal_market/consultations/2013/copyright-rules/index_en.htm and http://europa.eu/rapid/press-release_IP-13-1213_en.htm?locale=en

[161] See ' A single market for Intellectual Property Rights: COM (2011)287 final, 24/05/2011.

[162] See Implementation Appraisal 02/2014 Adapting the EU copyright rules to the digital transformation

[163] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2012:0789:FIN:EN:PDF

[164] See http://ec.europa.eu/internal_market/consultations/2013/copyright-rules/docs/consultation-document.pdf

the ongoing developments of the OpenScienceLink platform so we may have to take the new changes to the applicable legal framework into account when formulating the requirements in a later stage. This needs to be reviewed when the actual reform proposals are put into practice.

It is very likely that major changes will in particular be introduced to the existing system of limitations and exceptions to copyright and related rights in the Single Market. Currently, the InfoSoc Directive leaves much freedom to the Member States to introduce (or not) rules on exceptions in their national legislation[165], which in turn has significant implications on fair compensation for rights holders due to the cross border effect of limitations and exceptions. Changes that are considered include new exceptions relating to text and data mining (TDM) and maybe user generated content. It is furthermore contemplated to give the already existing exception for teaching a mandatory character. One important element that needs to be solved is to ensure cross-border effect of exceptions, in particular those that are relevant for internet transmissions of copyright protected content.

However, as long as copyright law is not fully unified in the EU (as is the case for trademark and design law), the territorial nature that characterizes copyright will remain a major hurdle for cross-border applications[166]. This territorial nature is very important for the OpenScienceLink project as it necessitates to take into account the differences that will continue to exist between the copyright laws of the various Member States. The project will clearly benefit from any reform that makes more copyright concepts uniform.

## 5.3 Copyright protection

To be able to determine if any of the works offered or used by OpenScienceLink falls under the scope of copyright we will first need to assess the 'work' requirements.

### 5.3.1 Literary and artistic works

The InfoSoc Directive does not provide a definition of the subject matter that is protected by copyright; hence, the general provisions of the Berne Convention find application. According to its article 2, a 'work' is protected by copyright if it constitutes "a literary or artistic work".[167]

This definition must be taken broadly and includes *every production in the literary, scientific and artistic domain*. It is very likely that most research data and scientific information considered by the OpenScienceLink Project will fall under the scope of copyright.

---

[165] See M-C. Janssens, "The issue of exceptions: Reshaping the keys to the gates in the territory of literary, musical and artistic creation", in Derclaye, E. (ed), *The Future of Copyright Law*, Series Research Handbooks in Intellectual Property, Cheltenham UK:Edward Elgar, 2009, p. 317.

[166] I. Hargreaves and B. Hugenholtz, "Copyright Reform for Growth and Jobs. Modernizing the European Copyright Framework", *Interactive Policy Brief*, Issue 13.2013; B. Hugenholtz, "Copyright Law without frontiers: the problem of territoriality in European Copyright Law", in in Derclaye, E. (ed), *The Future of Copyright Law*, Series Research Handbooks in Intellectual Property, Cheltenham UK:Edward Elgar, 2009, p. 12.

[167] The Berne convention states that member countries must *protect, in as effective and uniform a manner as possible, the rights of authors in their literary and artistic works* Given that most countries worldwide are part of the treaty the requirements for copyright protection have been mostly harmonized.

Excluded from protection are facts, ideas, systems, or methods of operation. [168] Unless these are expressed in whatever mode or form. The essence of copyright is indeed that protection is only afforded to *a perceptible expression*, not to the underlying ideas (expression-idea dichotomy)[169]. In addition, some countries have added the requirement that the work must be fixed in a tangible form to qualify for protection (e.g. the UK). Other formalities such as registration are not required (and may not be required!). [170]

Examples of copyrightable works include books, articles, drawings; photographic works; illustrations, maps, sketches and three-dimensional works relative to geography, topography, architecture or science. Collections of works can also be protected if they *by reason of the selection or arrangement of their contents*, constitute intellectual creations.[171]

The WIPO Copyright Treaty and the Agreement on Trade related aspects of intellectual property rights have both contributed to an expansion of the Berne framework. [172] Most notable is that the scope of works that fall under copyright protection now includes software and databases.

### 5.3.2 Requirement of Originality

The level of originality required within the meaning of the Berne Convention is today fully harmonised in the European Community.[173] A work only attracts copyright protection *'if it is original in the sense that it is its author's own intellectual creation'*. As regards computer programs, databases and photographs, explicit provisions are included in the relevant Directives. For all other works, harmonisation results from the case law of the European Court of Justice.*[174]*

The CJEU has further clarified that the originality of a work must be assessed through its components, for which the author had freedom in making creative choices.

[*I]t is only through the choice, sequence and combination […] that the author may express his creativity in an original manner and achieve a result which is an intellectual creation.* [175]

---

[168] Article 2 of the convention gives Member States freedom to decide if they further exclude official texts such as laws from copyright.

[169] See Article 9(2) TRIPs which states that "copyright protection shall extend to expressions and not to ideas, procedures, and methods of operation or mathematical concepts as such".

[170] Article 5(2) of the Berne Convention explicitly states that the enjoyment and the exercise of copyright shall not be subject to any formality. The granting of copyright on a work is automatic upon creation of that work.

[171] Article 2 (5) of the Berne Convention).

[172] Agreement on Trade related aspects of intellectual property rights (TRIPS) (1994).

[173] Van Eechoud, M. (2012). Along the road to uniformity - diverse readings of the Court of Justice Judgments on copyright work, *JIPITEC* 1, (60) 60-80 and Guibault, L., Licensing research data under Open Access conditions, in D. Beldiman (ed.), *Information and knowledge: 21st century Challenges in Intellectual property and knowledge governance,* Cheltenham, Edward Elgar, 2013.

[174] *'copyright… [within the meaning of Art. 2(a) InfoSoc] is liable to apply only in relation to a subject-matter which is original in the sense that it is its author's own intellectual creation'.* See case Infopacq, CJEU 16 July 2009 (C-5/08) and many subsequent cases.

[175] *Regarding the elements of such works covered by the protection, it should be observed that they consist of words which, considered in isolation, are not as such an intellectual creation of the author who employs them.*( case Infopacq, CJEU 16 July 2009 (C-5/08)).

*'An intellectual creation is an author's own if it reflects the author's personality'. That can be achieved if 'the author was able to express his creative abilities in the production of the work by making free and creative choices'.*[176]

### 5.3.3 Copyright protection for research data

If the research data available on the OpenScienceLink qualifies as an original work as defined above (*supra*, 5.3.1 and 5.3.2*)*, it will automatically be afforded copyright protection. This will have to be determined on a case to case basis and may be subject to interpretation by national courts. While in the past, significant differences could be perceived between Member States as regards the level of originality required for protection, recent case law of the CJEU has smoothed out most of these differences.

For the purpose of this paragraph we will distinguish between (1) data held in a database (2) a set of data and (3) a database. Because the EU directive on databases specifically deals with the protection of databases, this latter issue (3) will be analysed in the next chapter.

#### 5.3.3.1 Data

The subject matter of the data uploaded, shared and published on the OpenScienceLink platform is scientific data in the biomedical field. The sources from which OpenScienceLink collects its data are publicly available repositories that hold research data in the biomedical sciences. Most of the data will be factual. [177] For example sets of functional information, results of statistical analysis, descriptions of protein data, biological processes etc.

Facts and ideas are not protected by copyright unless they have been expressed in a tangible form with a sufficient level of originality. As most scientific research is about gathering information and discovering facts, the resulting individual data *as such* may not be considered copyrightable.[178] Hence, such data can be freely copied and re-used where accessible. We will refer to this data as raw data.

Examples of types of data and data sources within the OpenScienceLink platform[179] include:

[176] The CJEU has also made clear that, in principle, photographs can be copyrighted. In such case, choices can, for example, relate to pose, framing, angle, lighting and atmosphere, but also the use of developing techniques and 'post production' (Photoshop). 'By making those various choices, the author of a portrait photograph can stamp the work created with his "personal touch"'. CJEU 1 December 2011, case C-145/10 (*Eva Maria Painer*).

[177] Article 2 of the Berne convention in this respect mentions that (8) the protection of this Convention shall not apply to news of the day or to miscellaneous facts having the character of mere items of press information.

[178] This follows the general principles of scientific research where data should be recorded, handled, and stored in a way that allows accurate reporting, interpretation and verification. See the MRC Good Clinical Practice in Clinical Trials guidance. http://www.mrc.ac.uk/documents/pdf/good-clinical-practice-in-clinical-trials/

[179] For an overview of all the data sources currently included in the platform see D4.1 OpenScienceLink consortium p 23-25.

- The Gene Ontology which provides descriptions of the functional aspects of gene products. It covers the cellular component, parts of the cell and its environment and biological processes. This kind of data is not protected by copyright.[180]
- Clinical trails.gov contains key protocol details and summary results of clinical trials.[181] The data covers participant's flows, baseline patient characteristics, outcome measures and adverse events.[182]
- Kent Ridge Biomedical datasets which include gene expression data, protein filing data and genomic sequence data.
- The NIH Data sharing repositories, which include the Cancer Imaging Archive of medical images of cancer[183]. Some of these images may have attracted copyright if they are not purely technical. If the photographer sufficient freedom in selecting his subject, angle etc. Most medical images when they are merely technical are excluded from copyright protection but it may depend on a case to case basis whether the images fall under copyright protection.

#### 5.3.3.2 Datasets

Users can upload the underlying datasets of their publications to the platform. These datasets will include quantitative and qualitative research data such as survey results, clinical trial etc. Sets of scientific data are unlikely to attract copyright when they are the result of research (see, however, the possibility of database protection below). This can be concluded from the fact that the level of freedom required for a researcher *to express his creativity* is fairly limited. Paragraph 12 of the Declaration of Helsinki (DoH) requires that medical research involving human subjects must be justifiable on scientific grounds.[184]

However it is the researcher who *'Has to make sense of the data that have been collected by exploring and interpreting them'*.[185] When a researcher has sufficient freedom and made personal choices in how to present the results of his data collection activities the resulting dataset will be protected under copyright.

---

[180] This however does not mean that some data is not copyright protected. The Gene Ontology Consortium data and data products are licensed under the Creative Commons Attribution 4.0 Unported License. http://geneontology.org/faq-page

[181] Accessed at https://clinicaltrials.gov/

[182] In its terms and conditions it states that: 'The ClinicalTrials.gov data carry an international copyright outside the United States and its Territories or Possessions. Some ClinicalTrials.gov data may be subject to the copyright of third parties; you should consult these entities for any additional terms of use. These terms and conditions are in effect as long as the user retains any of the data obtained from ClinicalTrials.gov.'

[183] http://www.cancerimagingarchive.net/ which uses the CC BY 3.0 Unported licence see

[184] To ensure scientific merit, paragraph 12 of the DoH requires *a thorough knowledge of the literature on the topic and on previous laboratory and, where appropriate, animal research that gives good reason to expect that the proposed intervention will be efficacious in human beings. All research on animals must conform to ethical guidelines that minimize the number of animals used and prevent unnecessary pain*. Paragraph 16 adds a further requirement – that only scientifically qualified persons should conduct research on human subjects.

[185] In quantitative research, data analysis often only occurs after all or much of data have been collected. However, in qualitative research, data analysis often begins during, or immediately after, the first data are collected, although this process continues and is modified throughout the study.

#### 5.3.3.3 Articles in journals or other publications

The OpenScienceLink platform makes it possible for researchers to upload articles and (data) papers for peer review and OA Journal publishing. The platform services also enable linking to publications available outside of the OpenScienceLink platform.

In general, it is very likely that articles constitute an original writing by an author and will therefore fall under the scope of copyright protection. Artistic or other merit is not a requirement, so long as the article is the result of an intellectual creation that reflects the personality of the author (e.g. where the author has been able to choose his wording on how to describe his research results, conclusions and analysis).

Should a journal dictate a specific standard that does not leave any room for the author to make his own decisions with respect to the wording and information to include, the conclusion might be different. It seems, however, highly unlikely that no original input is apparent (the threshold of originality is generally typified as rather low). We refer to the *Infopaq* decision where the court ruled that even a mere sequence of 11 words in a newspaper article may enjoy copyright protection. [186] There is no reason to conclude differently for scientific articles. For most writings the level of originality required will be present.

#### 5.3.3.4 Metadata

'Metadata' are data that describe the characteristic features of certain information.[187] This can be a list of terms or keywords but may also include longer detailed descriptions. It is important for OpenScienceLink to ascertain when keywords and terms that constitute metadata fall under the scope of copyright.

As was said before, individual words as such are not copyrightable but (even short) combinations may attract copyright. In the Infopaq case the Court indeed confirmed that 'through the choice, sequence and combination of those words (…) the author may express his creativity in an original manner and achieve a result which is an intellectual creation'. The Court also held in particular that 'the possibility may not be ruled out that certain isolated sentences, or even certain parts of sentences in the text in question, may be suitable for conveying to the reader the originality of a publication such as a newspaper article, by communicating to that reader an element which is, in itself, the expression of the intellectual creation of the author of that article'.

[186] See para 44 'As regards newspaper articles, their author's own intellectual creation, referred to in paragraph 37 of this judgment, is evidenced clearly from the form, the manner in which the subject is presented and the linguistic expression. In the main proceedings, moreover, it is common ground that newspaper articles, as such, are literary works covered by Directive 2001/29'; CJEU July 2009, case C-5/08 (*Infopaq I*). See also CJEU 17 January 2012, case C-302/10 (*Infopaq II*).

[187] The metadata with a particular document may include such things as the name of the author, the number of pages, and the language in which the data is provided. The purpose of including metadata with the data to which it relates is to make the latter easier to find. Metadata is sometimes added automatically, but generally human intervention is required. KE-Cier report legal status of research data, p. 10.

The court furthermore addresses whether short extracts made by a third party without the authorization of the rights owner may constitute infringing reproductions under Article 2 Information Society Directive.

Hence, it will need to be assessed on a case by case base which metadata descriptions will – or will not be protected by copyright. We consider that, in the majority of cases, such will not be the case.

*Medical Subject Headings*

OpenScienceLink will make use of the Medical Subject Headings MeSH.[188] MeSH include three types of data namely (1) descriptors, (2) qualifiers and (3) supplementary concept records.[189]

- The descriptors or subject headings include short description or definition of the term.
- The qualifiers are subheadings to narrow down the descriptors.
- The concept records mainly describe chemical substances.

The purpose of these terms is to index scientific publications in the life sciences. For example the PubMed search engine uses MeSH as its main resource to index the biomedical scientific bibliography of the Medline repository. MeSH aims to provide a universally usable standard for index purposing. This and the fact that the types of data are not likely to leave any room for creative choice of words or definitions makes it unlikely that the MeSH metadata is copyright protected. This conclusion does not prejudice the possibility of collections of terms (see *infra*, protection of databases).

*Unified Protein Resource: protein sequences*

The Unified Protein Resource (Uniprot)[190] is a knowledge base which is the resource of information on protein sequences. It holds both the Swiss prot, which provides for manual and reviewed annotations to the sequences, and the TreMS which does this both the annotation and review automatically. The factual data itself: protein sequences are not protected by copyright. The annotations, if they only provide a standardised description, will normally not be protected either (e.g. the TreMS computerized annotations). A different conclusion may, however, be reached in the case of Swiss Prod which provides for a manually detailed analysis of the protein sequence and of the scientific literature; such expressions may benefit from copyright protection as the annotation may leave enough freedom in the choice of how to describe the sequences and relevant literature. [191]

*OpenScienceLink metadata provided by users to describe their uploaded datasets and publications.*

To index the articles and datasets and make them searchable both the articles and datasets have metadata attached. These include author, owner, license, data of publication, description, keywords, affiliation etc.[192] To make data retrievable most of this process follows a standardized procedure with commonly used terms and definitions. As regards possible copyright protection for such metadata, we refer to what

[188] MeSH terms are maintained by the United States National Library of medicine.

[189] See D4.1 OpenScienceLink platform Architecture and Background Platform Specification, p. 22.

[190] http://www.uniprot.org/

[191] The license for Uniprot reads Creative Commons Attribution-NoDerivs License to all copyrightable parts of our databases. However, if you intend to distribute a modified version of one of our databases, you must ask us for permission first.

[192] D4.1 Opensciencelink consortium, (2015)

was said in the beginning of this Section: while not obvious, copyright in metadata cannot entirely be ruled out and should be taken into account, especially when using metadata from other sources.

*Conclusion*

When Metadata refers to descriptions of biomedical data using standardised keywords and terms these are unlikely to be protected by copyright due to a lack of creative freedom in the choice, sequence and combination of the terms. The purpose of OpenScienceLink and the basis for its selection of data and services is the creation of a repository of well-structured and semantically linked datasets. The role of metadata herein is crucial and descriptions should be based on existing and established domain ontologies[193]. Any freedom in the description and definitions of terms may lead to less discoverability of the data and should therefore be avoided. As a result the amount of creative freedom for anyone to describe the data should be restricted also to avoid the issue of copyright ownership in metadata.

#### 5.3.3.5 Data repositories

The Database Directive defines a database as a collection of works or other materials that are systematically arranged and can be individually accessed.[194] The sources from which OpenScienceLink will gather its data such as the PubMed repository may qualify as a database if it fulfills the requirements..As regards the legal regime applicable to such data repositories – either by copyright and/or by the *sui generis* database right – we refer to our analysis in Section 5.6 ff.

#### 5.3.3.6 Interim conclusion on copyright protected works

Referring to the above mentioned analysis we can conclude that

- Not protected by copyright
  Raw data underlying the publications, datasets which lack originality and the metadata used to describe the data in a standardized way are not protected by copyright. To be copyrightable, the results of research must reach a level of creativity which is in general absent in scientific research results?
- Protected by copyright
  Research data expressed in publications, datasets and some metadata are likely to fall under copyright protection because and to the extent that - they meet the required threshold of originality.[195]

[193] D4.1 Opensciencelink consortium, (2015).
[194] Directive 96/9/EC, of the European Parliament and of the Council of March 11, 1996 on the Legal Protection of Databases, 1996 O.J. (L 77) 20.
[195] More specifically for the publication of scientific and scholarly works it depends on the copyright owner (1) whether OpenScienceLink can publish the work and (2) under which conditions other members of the scientific community and the general public can use the works. For the platform it depends whether OpenScienceLink (1) can use services such as data analysis to provide research trends and (2) include databases, data and datasets in its platform.

Since the OpenScienceLink platform services will make use of copyright protected works, one of the major consequences will be that prior authorization must be obtained for such services. Authorization can only be given by the owner of the copyright, which is not always the author himself.

### 5.3.4 Copyright ownership

In general the author who made the work will be granted first ownership. However since this issue is not harmonized, some national laws may provide for deviating rules in some situations.[196]

#### 5.3.4.1 Initial Author

On a European level, copyright regulations with respect to initial ownership of rights, have not been fully harmonized except in the case of software, databases and cinematographic works. This means that the requirements will depend on the national copyright regime of each Member State.[197]

For most countries however the person (author) who made the work is generally considered to be the initial author and thus rights owner.[198]

#### 5.3.4.2 Joint authorship

Where two or more authors have collaborated in creating a single work they may be considered as co-authors. Generally Member States require that consent is needed from all co-authors to alter or exploit the collaborative work. Hence, it is important for authors who wish to upload their publications to the OpenScienceLink platform, that they seek permission from all possible co-authors to make the work available on the platform.

#### 5.3.4.3 Works created in the course of employment

The general rule is that the author of the work is granted copyright. However this may be different in case of employment. Either because there is an explicit agreement between an employer and employee for the employer to obtain copyright or because the law states that in case of employment initial copyright is vested in, or transferred to the employer instead of the employee.

In the case of software programs, a transfer of rights is explicitly provided in the Software Directive. Only for such works, a fully harmonised solution exists in the EU. [199] In the case of database right a similar solution is proposed in the Database Directive but not imperatively imposed. Most Member States have, however, adopted this rule.[200]

---

[196] Goldstein P. (2001), International copyright: Principles, Law and Practice, Oxford University Press, pp. 25 and 205.

[197] Hugenholtz P.B., Van Eechoud M., Van Gompel S. et al. (2006), Recasting of Copyright and Related Rights for the Knowledge Economy, study prepared for the European Commission ETD/2005/IM/D1/95, Amsterdam. Retrieved at http://www.ivir.nl

[198] Goldstein P. (2001), International copyright: Principles, Law and Practice, Oxford University Press, pp. 25 and 205.

[199] Article 2 (3) of the Directive 2009/24/EC of the European Parliament and of the Council of 23 April 2009 on the legal protection of computer programs.

[200] See recital 29 and article 4 of the Database Directive.

It will be important to investigate, on a country-by-country basis, whether the rights of researchers to their research data and publications are – or are not – transferred to their employer, i.e. their research institution or university. In most European countries such a rule is not provided for publications but ownership of research data is often reserved by the institution. The OpenScienceLink platform will therefore need to make sure that the platform users who wish to upload their research data to the platform, verify that they are actually authorized to do so.

#### 5.3.4.4 Transfer of rights

The author of a work may wish to transfer all or some of his exclusive rights to others. This can be done by way of a license agreement or a full transfer of rights agreement. [201] As this matter is not harmonized, the conditions for there are differences between Member States with respect to the conditions under which copyrights such agreements will be deemed valid.

In common law countries (the UK and Ireland) no major obstacles for the transfer of rights exist. The civil law traditions follow two different approaches regarding the alienability of author's copyrights.

(1) The dualist approach, applicable in countries such as Belgium and France. The exclusive rights of the author include economic rights which can be transferred either through a licence or assignment. And moral rights which are inseparably linked to the author and cannot be transferred.
(2) The "monist" approach, which is applicable in countries such as Germany. Both the author's economic and moral rights are thoroughly intertwined. In practice, this means that authors in monist countries cannot assign any of their rights. They can only grant privileges to use the work (approximating nonexclusive and exclusive licenses in common law practice).[202]

Civil law countries have typically put in place mandatory rules to protect the author in his contractual relations with others. Therefore, as a rule, copyright contracts should be interpreted in a *restrictive* manner in favor of the author (*in dubio pro auctore*). Other conditions will relate to the instrument (requirement of a writing), the remuneration, the scope of the rights granted, the territory, etc.

With regards to publishers, often the author and publisher will enter into contractual agreements whereby the author will be required to transfer some or all rights.[203] There is, however, a growing tension between making research available as openly as possible, in an effort to expand the potential audience

[201] This is an important issue for any cross border use of works since it may also affect they validity of contracts signed with authors.
[202] Goldstein P. (2001), p. 219;
[203] Nature publishing (NPG) for example retains the copyright for commissioned articles. And for although it does not *require authors of original (primary) research papers to assign copyright of their published contributions.* It does require authors to grant NPG *an exclusive license to publish, in return for which they can reuse their papers in their future printed work without first requiring permission from the publisher of the journal.*

and increase potential citations, and publishers striving to retain their current subscription-based business models.[204] The open access policies are changing the traditional publishing model. [205]

In this respect, it is interesting to refer to a provision in the German Act[206] – and a comparable provision in a recent draft for an amendment of the Dutch Copyright Act[207] - that allows for the further dissemination of research results funded, in whole or in part, by public funds, upon the expiration of a certain period of time after first publication. It would be beneficial to the OpenScienceLink platform if other legislators were to follow this example.

## 5.4 Relevant acts and corresponding rights

### 5.4.1 Rights of the copyright owner

In the following section we will discuss the scope of the exclusive rights as well as exceptions and limitations as they are relevant for the OpenScienceLink platform. Where the former have been almost fully harmonized in the InfoSoc Directive, the issue of exceptions is still a very country-specific matter.

We will successively discuss the exclusive rights of reproduction (5.4.2.) and communication to the public (5.4.3), the one mandatory exception that allows for temporary reproductions (5.5.1) and other relevant exceptions (5.5.2).

### 5.4.2 The reproduction right in the InfoSoc Directive

Article 2 states that

*Member States shall provide for the exclusive right to authorise or prohibit direct or indirect, temporary or permanent reproduction by any means and in any form, in whole or in part:* [208]

*(a) For authors, of their works;*

The scope of the act of reproduction is very broad as it includes making 'temporary or permanent' copies ' by any means' in 'any form' in 'whole or in part' of works protected.

Recital 21 of the InfoSoc Directive provides that: *A broad definition of these acts is needed to ensure legal certainty within the internal market"*. This viewpoint is often repeated by the CJEU:

---

[204] Laakso, M. (2014). Green open access policies of scholarly journal publishers: a study of what, when, and where self-archiving is allowed. Scientometrics. In press. http://dx.doi.org/10.1007/s11192-013-1205-3.

[205] The SHERPA/RoMEO website provides a useful tool for researchers to find a summary of permissions that are normally given as part of each publisher's copyright transfer agreement. *Publisher copyright policies & self-archiving*. Available at http://www.sherpa.ac.uk/romeo/ .

[206] Section 38(4) Urheberrechtsgesetz.

[207] Proposed article 25 fa) of the Dutch Copyright Act, Amendment n° 8, introduced on 10 January 2015 (Parliamentary document n° 33 308): 'The creator of a short work of science for which the research is funded in whole or in part by public funds, has the right after a reasonable time after the first publication, to make the work available to the public for free, provided that the source of the first publication is indicated'.

[208] The various parts of a work enjoy protection under Article 2(a) of Directive 2001/29, provided that they contain elements which are the expression of the intellectual creation of the author of the work.

‘ [*a]n act occurring during a data capture process, which consists of storing an extract of a protected work comprising 11 words and printing out that extract, is such as to come within the concept of reproduction in part [..]If the elements thus reproduced are the expression of the intellectual creation of their author; it is for the national court to make this determination.* [209]

Some Member States have enacted separate provisions on the right to make adaptations or translations of the work, while others interpret these to be included in the definition of article 2.

For this section however all acts that involve making a copy of a copyright protected work include: the integration of research publications, abstracts, metadata and datasets from the platform users and external sources, elements of data analysis, the production of trend and evaluation reports if they include copyright protected elements and the storage and publishing of works online. OpenScienceLink will have to ask permission from all copyright owners involved to be authorized to perform all these acts, unless an exception applies. (Or the owner has provided a license)

The exclusive right of reproduction is subject to a mandatory exception that allows certain acts of temporary reproduction which will be discussed in the next Section.

### 5.4.3 The exclusive right of Communication to the public

Article 3 states that *1. Member States shall provide authors with the exclusive right to authorise or prohibit any communication to the public of their works, by wire or wireless means, including the making available to the public of their works in such a way that members of the public may access them from a place and at a time individually chosen by them.* [210]

*3. [T]he rights … shall not be exhausted by any act of communication to the public or making available to the public.*

As this text makes clear, the right of communication to the public includes a ‘making available right’.

To criteria to decide if there is a communication to the public are that there must be an “act of communication” of ‘a work’ and the communication of that work is directed ‘to a public’ which is either a “new public” or the same public but whereby a different means of transmission is used. Although the CJEU has had many occasions to clarify the scope of this communication right, some issues still remain unclear and more case law may be expected in the future that the OpenScienceLink project will need to take into account. From settled case-law, the following guidelines can be drawn:

- For there to be an “act of communication”, it is sufficient that a work is made available to a public in such a way that the persons forming that public may access it, irrespective of whether they avail themselves of that opportunity.
- There needs to be a public: “an indeterminate number of potential recipients and […] a fairly large number of persons”[211]

[209] CJEU 16 July 2009, C-5/08 (Infopaq).
[210] Paragraph 2 of Article 3 of the InfoSoc Directive, grants a comparable making available right in favour of owners of neighbouring rights (performers, producers and broadcasting organisations).
[211] CJEU 15 March 2012, Del Corso, C-135/10, paras 82 ff.

- which is a 'new' public that "was not taken into account by the copyright holders when they authorized the initial communication to the public" [212]

#### 5.4.3.1 Platform use

To encourage scientific research OpenScienceLink makes it possible for its users to use various types of research data including research results, publications but also raw data. Some data will fall under copyright protection and therefore use that is not covered by a statutory exception could be an infringement of the exclusive rights of the copyright owner. For any of the acts that fall under these exclusive rights permission is needed from the respective rights holder (unless an exception applies).

The relevant exclusive rights are:

*The right of reproduction*. The printing out of data, electronic copying on memory sticks and even the loading of the data into the working memory of a computer are acts of reproduction.

For OpenScienceLink this means that processing of the data in the electronic processing system and database is a relevant act of reproduction. The end user performs an act of reproduction by downloading a copy on his computer or making a print-out or screenshot to re-use in his own publications.

T*he right of distribution* remains somewhat unclear, however following the assessment of the OpenAIRE study: *if users are not being given a physical copy of the data on material storage media the right of distribution is not infringed.*[213]

*The right of communicating to the public* will not be involved unless use of protected content is made in such a way that it is also made available by the platform user through other channels.

#### 5.4.3.2 Linking and Embedding

For Open Access Data Journals the linking between research literature and primary datasets is important. The OpenScienceLink researcher is able to generate links between his research publications and associated datasets either stored in the OpenScienceLink repository or in external repositories and/or open data journals.

For many years, a debate has been going on relating to the question whether the use of hyperlinks amounts to a communication of a work and should be subject to the right holders' authorisation? The CJEU has addressed this question in the Svensson and the BestWater cases where it held that mere linking does not infringe the exclusive right "[..] if and to the extent that this work is freely accessible on the website to which the Internet link leads, there is an assumption that the copyright holders have, when they permitted this communication, considered all Internet users as the public". [214] Where a clickable link makes it possible for users of the site on which that link appears to circumvent restrictions

[212] CJEU 7 December 2006, SGAE/Rafael Hoteles C-306/05, para 40 and 42, CJEU 4 October 2011, C-403/08, Premier League and the order of 18 March 2010 in Case C-136/09 Organismos Sillogikis Diacheirisis Dimiourgon Theatrikon kai Optikoakoustikon Ergon, paragraph 38.

[213] This right was not discussed in the Section on exclusive rights because it does, as a matter of principle, not apply in an online environment.

[214] CJEU 13 February 2014, case C-466/12 (*Svensson)* and CJEU 21 October 2014, C-348/13 (*BestWater International*).

put in place by the site on which the protected work appears in order to restrict public access to that work to the latter site's subscribers only, it must thus be assumed that an infringing act takes place because the material will be accessible by a "new public". This will equally be the case where a hyperlink would be applied to content which has not been lawfully made available (hyperlinks to illegal source materials).

The court applied a same holding in the BestWater case in respect of 'embedding' technology: "… the embedding of a protected work which is publicly available on a website in another website by means of a link using the framing technology, does not by itself constitute communication to the public within the meaning of Article 3 (1) of Directive 2001/29, to the extent that the relevant work is neither communicated to a new public nor is it communicated using a specific technical means, which is different from that of the original communication".

OpenScienceLink will need to monitor that only publicly available information from lawful sources is being linked to or embedded within the platform. It also needs to make sure not to link to work which is no longer available to the public on the site on which it was initially communicated or if it is henceforth available on that site only to a restricted public. In those cases the copyright holders' authorization would be required for such a communication to the public.

## 5.5 Exceptions and limitations

### 5.5.1 Mandatory exception of temporary reproductions

The exception for temporary acts of reproduction is the only mandatory exception provided by the InfoSoc Directive.

Art. 5.1 reads as follows:

*Temporary acts of reproduction referred to in Article 2, which are transient or incidental [and] an integral and essential part of a technological process and whose sole purpose is to enable:*

*a. a transmission in a network between third parties by an intermediary, or*
*b. a lawful use*

*Of a work or other subject-matter to be made, and which have no independent economic significance, shall be exempted from the reproduction right provided for in Article* 2[215, 216]

215 Recital 27 emphasizes that providing communications facilities does not by itself qualify as making a communication. This exception must be implemented in full, and prevents coercion of Internet Service Providers running services such as Web caches with no independent value for their users. It also prevents rights holders from controlling all access to works through digital technologies, which by their very design make temporary "copies" of works as they are transferred from a medium such as a DVD to the player's memory for processing, and then to a display or speaker. Recital 33 of the InfoSoc Directive prevents proxies from altering cached data, and mandates "widely recognized and used" technology that provides information on access to data by users to rights holders. Given the minimization requirement of Article 6 of the Data Protection Directive, such technology should provide anonymized rather than personally identifiable information.

216 See also CJEU 4 Oct 2011, C-403/08 (Premier League) *(33) The exclusive right of reproduction should be subject to an exception to allow certain acts of temporary reproduction, which are transient or incidental*

The conditions are cumulative meaning that non-compliance with any one of them will lead to the act of reproduction not being exempted.[217] Because it is a restriction to the exclusive right of the author, the exception has to be interpreted strictly. There have been several court cases to further clarifying the scope of this exception. For OpenScienceLink, it is interesting to explore whether this exception could apply to certain acts of text and data mining which will be a core service of the OpenScienceLink platform. The following criteria will need to be met:

**Temporary, transient or incidental.** The analysis of the works during a data mining process may only involve temporary copies that are kept for the duration of the access to the work. It can involve transient or incidental copies.

"Incidental" in the definition means that the copy neither exists independently of, nor has a purpose independent of, the technological process of which it forms part.[218]An act can be held to be 'transient' only if its duration is limited to what is necessary for the proper completion of the technological process in question. [219]

The process of copying must be an automated process so that it deletes the copies automatically, without relying on human intervention.[220] The copies must be deleted as soon as their function of enabling the completion of the process has come to an end. [221]

---

*reproductions, forming an integral and essential part of a technological process and carried out for the sole purpose of enabling either efficient transmission in a network between third parties by an intermediary, or a lawful use of a work or other subject-matter to be made. The acts of reproduction concerned should have no separate economic value on their own. To the extent that they meet these conditions, this exception should include acts which enable browsing as well as acts of caching to take place, including those which enable transmission systems to function efficiently, provided that the intermediary does not modify the information and does not interfere with the lawful use of technology, widely recognised and used by industry, to obtain data on the use of the information. A use should be considered lawful where it is authorised by the right holder or not restricted by law.'*

[217] For an extensive study on this provision see DEPREEUW, S. and HUBIN, J.-B., "Study on the territoriality of the making available right. Localization of the act of making available to the public and its consequences" in TRIAILLE, J.-P. (ed.), "Study On The Application Of Directive 2001/29/EC On Copyright And Related Rights In The Information Society (The "InfoSoc Directive")", European Union, October 2013, p. 109 et s.,

[218] CJEU, 16 July 2009, C-5/08 (Infopaq I), para 43 ff.

[219] The relation between the transient and incidental copies was addressed in the case Meltwater where the Judge stated that: "If, as I consider, the copies made in the internet cache or on screen are "transient", it is strictly Speaking unnecessary to consider whether they are also "incidental". But I think it clear that they Are. The software puts a web-page on screen and into the cache for the purpose of enabling a Lawful use of the copyright material, i.e. viewing it. The creation of the copies is wholly incidental to the technological process involved" UK Supreme Court, 17 April 2013, [2013] UKSC 18.

[220] In the Infopaq-case, the CJEU concluded that the creation of files resulting from the conversion of files, is considered transient as long as they are deleted automatically from the computer memory. The storing of a text extract of only a few words is transient if this is automated with the result that that file is deleted promptly and without human intervention from the computer memory, and Printing out files containing the extracts of 11 words and thus reproduces those extracts on a paper medium is not transient.

[221] It was not disputed in the Infopaq-case that genuinely independent summary writing per se does not require consent from the right holders, but that the creation of a TIFF file when the printed articles are scanned and the conversion of the TIFF file into a text file is an act of reproduction and the storage and the printing out of the extract of 11 words.

**Technological process.** The copies must constitute an integral and essential part of those technological processes. *[B]e carried out entirely in the context of the implementation of the technological process and, not fully or partially, outside of such a process.* This is the case if the technological process of the services does not function correctly and efficiently without that act of reproduction of the data.[222]

**Sole purpose**.

*Transmission.* The temporary copies should "enable transmission systems to function efficiently, provided that the intermediary does not modify the information and does not interfere with the lawful use of technology, widely recognised and used by industry, to obtain data on the use of the information" [223]

*Lawful use.* The act of copying of sources could fulfil this condition if it has as sole purpose to enable a 'lawful use' of the works. If the intended purpose of the platform service process is to extract information (which may not be copyright protected) from various sources and derive new knowledge from them, then it could be qualified as lawful.[224] In Meltwater, the UK Supreme Court decided that acts of browsing (including the mere viewing, the access and the consultation of a webpage) constituted a lawful use justifying the making of transient copies generated by an end-user's use of the internet.[225] Also when a database is publicly available through a license that allows data analysis this will be considered a lawful use. If the right holder expressly reserved his rights in terms and conditions, the extraction of its data will be unlawful.

**No independent economic significance.** According to the Court of Justice, there is an independent economic significance if the author of the reproduction is likely to make a profit due to the economic exploitation of the temporary reproduction itself or if the act of temporary reproduction leads to a change in the subject matter reproduced. Such act no longer aims to facilitate the use of the work, but the use of a different subject matter.[226] Important for OpenScienceLink in the data analysis for making trend reports is that in order to comply with the condition of not having an independent economic significance, the acts of temporary reproduction may not lead to a modification of the work.[227]

A recent EU study on Text and Data mining concluded that it is rare that services like the ones that are being developed within the OpenScienceLink platform for text and data mining can benefit from the exception of Article 5.1 of the InfoSoc Directive. [228] Because they have analysed many steps within the process including

---

[222] See, in particular, CJEU 5 June 2014, case C-360/13 (*PRCA/NLA*).

[223] Recital 33 InfoSoc Directive.

[224] See in Infopaq II, the process which permits a more efficient drafting of summaries of newspapers articles was not considered unlawful. In Premier League, the picking up of broadcasts in private circles was judged as having the sole purpose of enabling a lawful use of the works.

[225] UK Supreme Court, 17 April 2013, [2013] UKSC 18.

[226] CJEU, 17 January 2012, Case C-302/10, Infopaq II, par. 51-53.

[227] CJEU, 17 January 2012, Case C-302/10, Infopaq II, par. 54.

[228] It seems that only in a few exceptional cases, copies involved in the steps of a mining process may fulfil the conditions of the temporary copy; the copies being transient or incidental (for a duration limited to what is necessary for the proper completion of the technological process), being part of a technological process (a mining process), having as sole purpose to enable a lawful use of the work (the only copy of ideas of works, which is not

- The obtaining of sources;
- The transformation of data to fit operational needs;
- The loading of data;
- The analysis of data, and
- The drafting of reports.

A detailed analysis of which platform services may be covered by this exception goes beyond the scope of this report. We will however make an analysis of the relevant licenses of the publicly available data sources to see if the specific use falls within the granted permission.[229]

### 5.5.2 Non- mandatory Exceptions and limitations

The InfoSoc directive includes in its Article 5 a long list of exceptions and limitations to the harmonized rights of reproduction and communication to the public. Except for the exception of *temporary reproductions*, discussed above, all other exceptions are optional which means that national legislators remain free to (not) include them in their national copyright acts and/or to limit (not extend!) their scope of application. The list itself is, however, exhaustive[230], and Member States are not allowed to design other exceptions. To ensure compliance with international obligations[231] all exceptions and limitations must moreover comply with the three-step test of Article 5(5) InfoSoc Directive. Thus, exceptions can only be applied in certain special cases which do not conflict with a normal exploitation of the work or other subject-matter and do not unreasonably prejudice the legitimate interests of the right holder.

The optional exceptions in Article 5(2) and (3) relate to diverse purposes and include

*private copying; use of copyrighted material by libraries, museums and archives; ephemeral recordings; reproductions of broadcasts made by hospitals and prisons; illustrations for teaching; use for scientific research; use for the benefit of people with a disability; press privileges; use for the purpose of quotations, caricature, parody and pastiche; use for the purposes of public security and for the proper performance or reporting of administrative, parliamentary or judicial proceedings; use of political speeches and public lectures; use during religious or official celebrations; use of architectural works located permanently in public places; incidental inclusions of a work in other material; use for the purpose of advertising the public exhibition or sale of artistic works; use in connection with the demonstration or repair of equipment; use for the reconstruction of buildings; and additional cases of use having minor importance.*

restricted by copyright, or another use allowed by the right holder or another exception) and finally if the copies have no independent economic significance, separable from the economic advantage derived from the lawful use of the work concerned or do not lead to a modification of that work.

[229] This analysis is part of the upcoming revision of the further developments within the OpenScienceLink platform.

[230] Recital 32 of the InfoSoc Directive refers to an 'exhaustive enumeration of exceptions and limitations to the right of reproduction and the right of communication to the public'.

[231] See Article 9(2) of the Berne Convention for the Protection of Literary and Artistic Works (1971); Article 13 of the TRIPS Agreement (Trade Related Intellectual Property Rights) 1994;and Article 10 of the WIPO Copyright Treaty (1996).

#### 5.5.2.1 Exception for scientific research (Art. 5.3. a InfoSoc Directive)

Article 5.3 of the InfoSoc directive provides for an exception to the right of reproduction and to the right of communication to the public when the protected work is used*: "for the sole purpose of illustration for teaching or scientific research, as long as the source, including the author's name, is indicated, unless this turns out to be impossible and to the extent justified by the non-commercial purpose to be achieved"*.

The optional exception for scientific research is considered to be an important exception to the exclusive rights, for achieving open access to scientific data.[232] However not all Member States have transposed this exception whereas some have implemented it in a detailed way and others have national legislation that is more restrictive and narrower than the Directive[233] This makes it problematic for OpenScienceLink to be able to rely on this exceptions.

The cumulative conditions under the InfoSoc Directive are [234]/[235]

1. works are used for the sole purpose of illustration for teaching or scientific research;[236]
2. the source, including the author's name, is indicated, unless this turns out to be impossible;
3. Works are used to the extent justified by the non-commercial purpose to be achieved. Furthermore, the exception can only be applied to the extent that
4. The use of the works does not conflict with a normal exploitation of the work or other subject-matter and does not unreasonably prejudice the legitimate interests of the right holder (three-step test).

It should be noted that the European Commission is currently considering to make this provision mandatory. Should this occur, it will have to been examined what the possible impact for the OpenScienceLink platform will be.

#### 5.5.2.2 Exception of private copying (art. 5.2.b InfoSoc Directive

Article 5(2) b of the InfoSoc Directive allows Member States to provide an exception that permits the making of "reproductions on any medium made by a natural person for private use and for ends that are neither directly nor indirectly commercial, on condition that the right holders receive fair compensation

[232] Study on the legal framework of text and data mining, De Wolf & Partners, 2014.

[233] It was concluded that the main differences between the Member States: (a) the diverging objectives pursued by the national provisions, (b) the beneficiaries and the users, (c) the works concerned, (d) the authorized acts, and (e) the implementation of the other conditions contained in Article 5(3) of the InfoSoc Directive. See DUSOLLIER, S., "The Limitations and Exceptions to Copyright and Related Rights for Libraries, Research and Teaching Uses", in TRIAILLE, J.-P. (ed.), "Study On The Application Of Directive 2001/29/EC On Copyright And Related Rights In The Information Society (The "InfoSoc Directive")", European Union, October 2013, p. 109 ff.

[234] As was said before, Member States may limit the scope of application. The text of the Directive contains the broadest possible form of use.

[235] See for a detailed analysis (Study on the legal framework of text and data mining, De Wolf & Partners , 2014)

[236] According to the Explanatory Memorandum, paragraph 36: "the term 'scientific research` within the meaning of this Directive covers both the natural sciences and the human sciences".

which takes account of the application or non-application of technological measures referred to in Article 6 to the work or subject-matter concerned".[237]

The private copying exception on the exclusive right of reproduction allows users to reproduce works when this is strictly for private purposes. Therefore, commercial use of reproductions, as well as communication to the public, distribution to the public, public performance or adaptation is out of the scope of private copying. [238] Content retrieved from illegal sources does not fall within the scope of private copying either. [239] Where private use is allowed, significant differences among Member States remain as regards the obligation to pay fair compensation. [240]

In general the conditions to fall under the exception include that

(1) The person is an individual,

(2) The copy has been acquired legally

(3) The further copy is for the individual's private use

(4) The further copy is made for non-commercial purposes.

Moreover, in view of article 5.5 of the InfoSoc Directive, the exception shall only be applied cases which do not conflict with a normal exploitation of the work or other subject-matter and do not unreasonably prejudice the legitimate interests of the right holder. All these conditions have to be assessed by courts in individual cases.

### 5.5.3 Conclusion

The current framework of exceptions and limitations does not offer significant solutions for the uses that are intended by the OpenScienceLink platform. As a consequence, it will be important for OpenScienceLink to 'clear' the necessary rights to enable it to make research data available in open access. It should prior to such use, seek the authorization from the respective copyright holders. Given the differences in national law, it will be important to impose the responsibility for rights clearance with the user of the platform who wishes to upload a work. With respect to the acts performed by OpenScienceLink itself, rights clearance needs to take place before including any of the data from external datasets and third party rights holders.

---

[237] The right to private copying was clarified by the CJEU in Padawan SL v Sociedad General de Autores y Editores de España (SGAE) (C-467/08) 2010, where the court stated that "Natural persons are rightly presumed to benefit fully from the making available of that equipment, that is to say that they are deemed to take full advantage of the functions associated with that equipment, including copying."

[238] See http://ec.europa.eu/internal_market/copyright/docs/studies/140623-limitations-economic-impacts-study_en.pdf, p. 82.

[239] CJEU, case C-435/12 ACI Adam BV and Others v Stichting de Thuiskopie

[240] A. Vitorino (2013) Recommendations resulting from the 'mediation on private copying and reprography levies. http://ec.europa.eu/internal_market/copyright/docs/levy_reform/130131_levies-vitorino-recommendations_en.pdf.

## 5.6 The database directive

The Directive on the legal protection of Databases was adopted in February 1996. The Directive created a new exclusive "*sui generis*" right for database producers. Furthermore it harmonized copyright for the structure and arrangement of the contents of databases ("original" databases). [241]

This protection is only available for databases with comply with the definition in Article 1.2 of the Database Directive, i.e. "a collection of independent works, data or other materials arranged in a systematic or methodical way and individually accessible by electronic or other means".[242] This excludes works such as books, midi files or videogames which may include various types of materials (words, images) but these elements are not individually accessible.

Recital 17 of the Database Directive further states that: "(…) the term 'database` should be understood to include literary, artistic, musical or other collections of works or collections of other material such as texts, sound, images, numbers, facts, and data".[243]

## 5.7 Copyright protection for databases

### 5.7.1 Originality

To be given copyright protection, a database must be 'original', i.e. the author's 'own intellectual creation' by reason of the selection or arrangement of the contents. As a consequence of the harmonization of the originality test for all works (see previous Section 5.3.), "sweat of the brow" databases which are not creative but are based on a certain level of effort or investment, do no longer qualify for copyright protection.[244]

In the Football Dataco case, the CJEU stated that: *the fact that the setting up of the database required, irrespective of the creation of the data which it contains, significant labour and skill of its author, […], cannot as such justify the protection of it by copyright under Directive 96/9, if that labour and that skill do not express any* ***originality in the selection or arrangement of that data****.*[245]

A database selection or arrangement, which purely depends on technical factors or aims to achieve accuracy and exhaustiveness, will therefore not be protected by copyright. Hence, many scientific databases will generally not be protected by copyright. This was concluded in a recent study on the

---

[241] The Directive is not applicable to a database which is not protected either by copyright or by the *sui generis* right under that directive, so that Articles 6(1), 8 and 15 of that directive do not preclude the author of such a database from laying down contractual limitations on its use by third parties, without prejudice to the applicable national law. CJEU 15 Jan 2015, C-30/14 (Ryanair).

[242] The term 'database' refers to any collection of works, data or other materials, separable from one another without the value of their contents being affected, including a method or system of some sort for the retrieval of each of its constituent materials. Judgment of 9 Nov 2004, C-444/02 (Fixtures Marketing).

[243] The CJEU has stressed the broad definition of "database" in the Directive See Case C-444/02 (Fixtures Marketing Ltd v. Organismos prognostikon agonon podosfairou AE - "OPAP"), n. 20, 25. And see (Study on the legal framework of text and data mining, De Wolf & Partners , 2014).

[244] http://ec.europa.eu/internal_market/copyright/docs/wipo/wipo-protection-db_en.pdf

[245] Case C-604/10 *Football Dataco v. Yahoo UK Ltd.*

protection of research data.[246] The reason is that there is no room for the author to exercise creativity or originality in the choice, sequence and combination of data in the collection. [247]

### 5.7.2 Interim conclusion

The goal of the OpenScienceLink database is to provide researchers access to data from as many publicly available sources of research data in the biomedical field as possible, it does not replace any existing repositories but does aim at indexing all the research data available. This aim leaves little room for 'creative' choice of what data should be included. Hence it seems unlikely, although not entirely excluded, that the collections of research data that are used within the OpenScienceLink platform as well as the OpenScienceLink platform database itself will meet the criterion of 'originality'. In case of a dispute, the competent national Court will have to resolve this discussion.

### *5.7.3* Exclusive Rights

Article 5 of the Database Directive states that with respect to an original database: 'the author of the work shall have the exclusive right to carry out or authorise:

a) Temporary or permanent reproduction by any means, in any form, in whole or in part;
b) Rights of adaptation, translation, arrangement and any other alteration;
c) Any form of distribution to the public of the database or of copies thereof (subject to Community exhaustion); and
d) Any communication to the public, display or performance to the public;
e) Any reproduction, distribution, communication, display or performance to the public of the results of the acts referred to in (b).

It is generally considered that these exclusive rights must be read in relation with the InfoSoc directive.[248]

### 5.7.4 Exceptions

Article 6 of the Directive lists the exceptions to the exclusive rights applicable to copyright protected databases. This includes the obligation to implement the lawful use exception. Furthermore the Member States have the option to provide for limitations on the exclusive rights:

> *(a) In the case of reproduction for private purposes of a non-electronic database;*
>
> *(b) Where there is use for the sole purpose of illustration for teaching or scientific research, as long as the source is indicated and to the extent justified by the non-commercial purpose to be achieved;*

---

[246] Lucie Guibault and Andreas Wiebe (Eds.) *Safe to be open, Study on the protection of research data and recommendations for access and usage,* OpenAIRE Project, Universitätsverlag Göttingen, 2013.

[247] Regarding scientific databases, the OpenAIRE report states that these are intended to be as complete as possible and the arrangement of data is usually a standard scientific practice. Therefore, these databases will regularly fail to meet the requirement of an intellectual creation and thereby be very rarely protected as database works under copyright law. Available at https://www.openaire.eu/

[248] DG INTERNAL MARKET AND SERVICES WORKING PAPER First evaluation of Directive 96/9/EC on the legal protection of databases.

*(c) Where there is use for the purposes of public security of for the purposes of an administrative or judicial procedure;*

*(d) Where other exceptions to copyright which are traditionally authorized under national law are involved, without prejudice to points (a), (b) and (c).*

As long as any of the exceptions do not allow 'its application to be used in a manner which unreasonably prejudices the right holder's legitimate interests or conflicts with normal exploitation of the database'.

#### 5.7.4.1 Exception for normal use

The legitimate interests of lawful users are safeguarded by the mandatory exception which permits all acts, which are necessary to obtain access to the contents of the database and normal use of the contents by the lawful user

*[T]he performance by the lawful user of a database or of a copy thereof … which is necessary for the purposes of access to the contents of the databases and normal use of the contents by the lawful user shall not require the authorization of the author of the database. Where the lawful user is authorized to use only part of the database, this provision shall apply only to that part.* [249]

This exception applies to the entire database as well as to parts of the contents of the database. The lawful user may continue to perform certain acts necessary to access the contents of databases and facilitate the dissemination of information.

- Lawful user

The Directive does not provide a definition of "lawful user". Recital 34 of the Database Directive refers to a user, who is authorised by agreement with the right holder, to access and use the database.

The authors of the EU study on TDM concluded that the majority of the users of a database protected by the *sui generis* right will be lawful users. [250] They define the lawful user as a *user who can invoke either a contractual authorization to use the database or a legal exception, provided that (1) in the latter case, no agreement which he has consented to prevents the application of the exception and that (2), in all cases, he did not circumvent any technical system meant to limit or prevent his use.* The authors further considered that *where no contractual conditions have been imposed and a website is freely accessible, we do not think there exists an implied license, but rather that no license is needed in the first place.* [251]

- Normal use of the contents

"Once the right holder has chosen to make available a copy of the database to a user, whether by an on-line service or by other means of distribution, that lawful user must be able to access and use the database

[249] Articles 6.1. of the Database Directive.
[250] 'The unlawful user will have had to steal or gain access to a stolen or infringing copy of a database; see the Study on the legal framework of text and data mining, De Wolf & Partners (2014);
[251] *Ibid.*

for the purposes and in the way set out in the agreement with the right holder, even if such access and use necessitate performance of otherwise restricted acts".[252]

Since access to and use of a database normally entails acts of reproduction, this exception primarily concerns the reproduction right.[253] The act must, however, be *necessary* for the purposes of accessing the database and the normal use. This may, e.g., include acts of searching (querying) and browsing on-line database and downloading the results of a search.[254]

It should be remembered that the rights holder may limit the access to his database and set forth conditions for the use of a database. However it is important to note that the exception as a whole cannot be waived by contract.[255]

Relevant for OpenScienceLink is thus to closely examine whether the licenses that apply to databases that are used, allow for the contemplated activities. It can either be that automatic accessing the database for example by data mining software is not allowed or the use for data mining purposes does not fall under what the database owner states as normal use. In these cases OpenScienceLink needs permission from the holder of the *sui generis* right to perform its platform services.

#### 5.7.4.2 Exceptions for Scientific research

The exception for scientific research[256] in relation to databases contained in Article 6(2) (b) of the Database Directive is optional in nature and has not been implemented in all Member States.[257] This again makes it problematic for OpenScienceLink to rely on this exception for its platform services.

In general the use which is not normal use of the structure of an original database will not infringe the authors' rights when:

(i) The database is used for the sole purpose of scientific research;
(ii) The sources are indicated;
(iii) The database is used to the extent justified by the non-commercial purpose to be achieved.[258]

It remains to be seen if OpenScienceLink can indeed rely on the exception given that its sole purpose is not scientific research as such. OpenScienceLink itself is not a research institute but only develops services for (scientific) users. A further analysis of the Member States' interpretation of this exception

[252] Recital (34) Database Directive.
[253] B. Hugenholtz, "Directive 96/9/EC (Database Directive), in Th. Dreier and B. Hugenholtz, *Concise European Copyright Law*, Kluwer Law Int. 2006, p. 324.
[254] B. Hugenholtz, "Directive 96/9/EC (Database Directive), in Th. Dreier and B. Hugenholtz, *Concise European Copyright Law*, Kluwer Law Int. 2006, p. 324.
[255] see Article 15 of the Database Directive.
[256] Recital (36) explains that the term scientific research covers both the natural sciences and the human sciences.
[257] The conditions of Article 6(2) of the Database Directive are similar to the conditions of Article 5(3)(a) of the InfoSoc Directive.
[258] The three-step test also applies as condition (article 6.3. of the Database Directive).

will be necessary to determine whether the OpenScienceLink services can benefit from this exception, although this seems not likely.[259]

#### 5.7.4.3 Conclusion

Considering the platform services of OpenScienceLink we see that if any of the databases fall under copyright protection they will require authorization. But only if the use does not fall under the broad exception of what constitutes normal use by a lawful user. With respect to the optional exceptions, the lack of harmonization creates legal uncertainty as to the possibility for OpenScienceLink to benefit from any of them.

It should be emphasized that the protection that is offered through the copyright regime only concerns the (original) structure of databases. It seems unlikely that OpenScienceLink will reproduce or use this structure in any manner. More important for OpenScienceLink is the question to what extent it can make use of contents of third party databases. Retrieving such content may be more problematic from the perspective of another right that protects databases, namely the so-called *sui generis right*, discussed in the next Section.

## 5.8 Sui generis right ('SGDR')

The *sui generis* database right is specifically intended to safeguard the position of makers of databases against misappropriation of the results of their investments of financial resources time, effort and energy. [260] The *sui generis* right has thus a very specific objective that is not to protect a creation (as is the case with most intellectual property rights) but to protect an investment in producing a database.[261]

When enacted in 1996, the Database Directive introduced 'new law' for most Member States in respect of non-original databases.[262] This explains why most national legislators have opted for a rather faithful reproduction of the provisions of the Directive, as, prior to its publication, database producers were not granted any exclusive right to protect their databases. It can therefore be assumed that what is explained in this Section, is common in all Member States.

[259] The authors of the EU study on text and datamining concluded that services such as data analysis fall outside of the scope of the exception from the moment it is also done for other purposes than scientific research; for example for statistical or behavioural analysis. see Study De Wolf & Partners (2014).

[260] According to recital 7 of the Directive, the *sui generis* right was developed because "the making of databases [requires] the investment of considerable human, technical and financial resources while such databases can be copied or accessed at a fraction of the cost needed to design them independently."

[261] M-C. Janssens, "Implementation of the 1996 Database Directive into Belgian Law", *IIC – International Review of Industrial Property and Copyright Law* (D.), vol. 31, 2000/1, p. 52; P.B. Hugenholtz, " Implementing the European Database Directive", in *Intellectual Property and Information Law. Essays in Honour of Herman Cohen Jehoram*, Kluwer, 1998, p. 183.

[262] B. Hugenholtz, "Directive 96/9/EC (Database Directive), in Th. Dreier and B. Hugenholtz, *Concise European Copyright Law*, Kluwer Law Int. 2006, p. 307 and 327.

The new protection scheme is independent from – and does prejudice upon - the copyright protection of the database, as well as from any copyright to elements of its content[263]. Hence, a database can at the same time enjoy the protection of copyright as well as the *sui generis* right.

### 5.8.1 Substantial investment

The main prerequisite for a collection of separate items of data to be protected as a database, is that there has been a *qualitatively and/or quantitatively substantial investment.* Such investment must be done *in either the obtaining, verification or presentation of the contents of the database*. [264] The Directive itself does not provide many indications as to when a database involves a *substantial* investment.[265]

Clarification has been given by the CJEU in several landmark cases.[266] The court ruled that:

- an investment in the *creation* of data does not fall within the notion of 'obtaining' and should therefore be disregarded. 'Obtaining' refers to the gathering of pre-existing data or materials

- an investment in the 'verification of the contents' refers to the resources used, with a view to ensuring the reliability of the information contained in the database and/to monitor the accuracy of the materials collected; it thus involves the checking, correcting and updating of data that already exists in the database[267]

- an investment in the 'presentation of the contents' concerns activities related to the retrieval and communication of the contents of the database, such as digitalizing analog files, producing a thesaurus or designing a user interface.[268]

---

[263] Football staccato states in 14 that protection under the copyright focuses essentially on the structure of the database, that is, the way in which it has actually been put together through the selection of the data to be included or the way in which they are presented. What is more, Article 3(2) states clearly that the copyright provided for in that article 'shall not extend to [the] contents' of databases, which can be protected by copyright autonomously, but are not protected by virtue of being entered in a protected database. Recital 15 to the Directive states that the copyright protection 'cover[s] the structure of the database'. The '*sui generis*' protection, on the other hand, is simply a right to prohibit extraction and/or re-utilisation of the data contained in the database. That right is conferred, not to protect the originality of the database in itself, but to compensate the effort expended in obtaining, verifying and/or presenting the data contained therein (Case C-46/02 Fixtures Marketing, para 39).

[264] Article 7 Database Directive. For more details on this condition, see CJEU, C-203/02 *British Horse Racing Board ltd. And others v. William Hill Organization ltd.*

[265] According to VANHEES, the development and commercialization of a database demands a considerable investment if: (1) data has to be collected; (2) permission has to be obtained for their storage and use in the database; (3) the collected data has to be selected, processed and organized; (4) instruments have to be made to find separate elements of the database; and (5) the contents of the database have to be verified and updated Vanhees, H. "De juridische bescherming van databanken" (2001), *De juridische bescherming van databanken,* Antwerpen, Kluwer, p. 6 and Janssen, K. & Dumortier, J. (2006)

[266] See the cases Fixtures Marketing Ltd v Svenska AB (Svenska), C-338/02; Fixtures Marketing Ltd v Organismos Prognostikon Agonon Podosfairou EG (OPAP), C-444/02; Fixtures Marketing Ltd v Oy Veikkaus Ab (OyVeikkaus), C-46/02 and British Horseracing Board Ltd v William Hill Organization Ltd (BHB decision), C-203/02.

[267] B. Hugenholtz, "Directive 96/9/EC (Database Directive), in Th. Dreier and B. Hugenholtz, *Concise European Copyright Law*, Kluwer Law Int. 2006, p. 329.

[268] *Ibid.*, at p. 329.

The application of these requirements to scientific databases will not always be easy. In particular, the distinction to be made between the investments into the creation of data (not protected) and to obtain the data (protected) is cumbersome. Davis and Hugenholtz put it as follows:

"*One example may be the recording of meteorological data such as the daily maximum temperature in a particular location. Are those data created or obtained? Similarly, do scientists obtain the genetic sequences of living organisms or do they create them? The strict approach taken by the ECJ in these four cases would suggest that the answer is that such data are created. Meteorological data and genetic sequences are records and representations of natural phenomena, not the phenomena themselves, and it would be difficult for scientists to argue that they have simply collected the data as opposed to creating them. On the other hand, when a large mass of such data has been created, there are also significant costs associated with presentation and verification which may meet the requirements in Article 7(1) of the Directive. In any event, these metaphysical distinctions will undoubtedly continue to concern courts, and commentators for some time to come".*[269]
But there are also authors that defend the opposite view.[270] This legal uncertainty is undoubtedly problematic for the OpenScienceLink project.

### 5.8.2 Database Rights holder

One of the consequences of the goal of the *sui generis* right to protect investments into database creation is that the right does not originate with the maker of the database but belongs - *ab initio* - to the (legal or natural) person that makes the investment.[271] Recital 41 defines the 'maker of the database' as the person who takes the initiative and the risk of investing, excluding sub-contractors.[272] The investment can be through financial resources and/or through time, effort and energy.

It is important to note here that the right holder does not in any way 'own' the facts or bare data.[273] A right holder cannot prevent the making of a similar or identical database by a competitor who independently with money time and effort creates a similar or identical result.

When several databases are combined to form the OpenScienceLink platform, permission will be needed from the owners of the *sui generis right* that may possibly – if the conditions for protection are met – be

[269] M. Davison & P. Bernt Hugenholtz, "Football fixtures, horse races and spin-offs: the ECJ domesticates the database right", *EIPR* 2005, p.113.
[270] E Derclaye 'Database "*Sui generis*" Right: Should we Adopt the Spin-off Theory' *EIPR*, 2014, p. 402; D. Visser" A Database from Space, the Legal Protection of Data Created or Collected in Outer Space under the 1996 European Database Directive", in, *The International Space Station: Commercial Utilisation from a European*, Martinus Nijhoff Publ., 2006, p. 101-102.
[271] See article 7 Database Directive. See also Recital 41 of the Database Directive ('the person who takes the initiative and the risk of investing').
[272] See, e.g., a Dutch *Landmark*-case, where the city of Amsterdam was held the producer of the database, because the city had (1) invested substantially in the creation of this database, (2) it was commissioned by the city, (3) the city took all the risk and (4) the vast majority of the money invested came from the Minister in the form of a subsidy. See Raad van State, 29 April 2009, *Gemeente Amsterdam v Landmark.* Retrieved at http://www.boek9.nl/files/2009/IEPT20090429_RvSt_Gemeente_Amsterdam_v_Landmark.pdf and De Vries, M., Reassessing the 2009 Landmark case, *EpsiPlatform*, 2011 at http://epsiplatform.eu/content/reassessing-2009-landmark-decision
[273] Recital 45 of the Database Directive.

attached to such databases. As regards the new database that will be created, it may or may not constitute a protected database, depending on whether the database involved a 'substantial investment' into either the obtaining, verification or presentation of its contents.

### 5.8.3 Term of protection

The database right lasts for a term of 15 years from the date of completion of the database.[274] This term can be extended – without limit! – upon each additional substantial investment in the contents of the database.

Like copyright protection, the *sui generis* database right arises automatically, without any formal requirement, at the moment the database is completed or disclosed to the public.

### 5.8.4 Scope of protection

The *sui generis* right is an exclusive right that comprises two distinct attributes; the owner of the right can prevent the *extraction* and/or *re-utilization* of the *whole or of a substantial part* of the contents of that database.[275] It is important to note that this right does not include the possibility to prevent the use of non-substantial parts of a database: these acts remain free to all. No definition is given of what constitutes ‘a substantial part’ and this decision will have to be taken by individual courts on the bases of the particular facts of the case. The Directive does indicate in this respect that substantiality has to be assessed either on a qualitative (value of the data) or a quantitative (amounts of data) basis (art. 7).

In 2004 The CJEU has provided the following clarification as regards a ‘substantial’ part evaluated quantitatively or qualitatively.[276]

§69.[…] whether the part at issue is substantial, must, refer to the investment in the creation of the database and the prejudice caused to that investment by the act of extracting or reutilising that part.

§70. The expression 'substantial part, evaluated quantitatively', of the contents of a database within the meaning of Article 7(1) of the directive refers to *the volume of data* extracted from the database and/or re-utilized, and must be assessed in relation to the volume of the contents of the whole of that database. If a user extracts and/or re-utilises a quantitatively significant part of the contents of a database whose creation required the deployment of substantial resources, the investment in the extracted or re-utilised part is, proportionately, equally substantial.

§71. The expression 'substantial part, evaluated qualitatively', of the contents of a database refers *to the scale of the investment* in the obtaining, verification or presentation of the contents of the subject of the act of extraction and/or reutilisation, regardless of whether that subject represents a quantitatively substantial part of the general contents of the protected database. A quantitatively negligible part of the contents of a database may in fact represent, in terms of obtaining, verification or presentation, significant human, technical or financial investment. […]

[274] Article 10 Database Directive.
[275] Article 7 Database Directive.
[276] British Horseracing Board Ltd v William Hill Organization Ltd (BHB decision), C-203/02.

#### 5.8.4.1 The exclusive right of 'extraction',

The right of extraction is defined as "the permanent or temporary transfer of all or a substantial part of the contents of a database to another medium by any means or in any form". The right includes the downloading, copying, printing, or any other reproduction in whatever (permanent or temporary) form. Important to note here is the fact that also 'temporary' reproductions fall under the exclusive rights of the database-rights holder.

The CJEU has, however, decided that the mere consultation of a database, is not covered by the *sui generis* right.[277]

#### 5.8.4.2 The exclusive right of re-utilization'

The database-rights holder also has **t**he right of re-utilization. This is defined in Article 7 as "any form of making available to the public all or a substantial part of the contents of a database by the distribution of copies, by renting, by on-line or other forms of transmission".

#### 5.8.4.3 Extraction of "insubstantial parts" by the lawful user

Although this results from the definition (see *supra*, 5.8.1), article 8.1 of the Database Directive confirms the following: "The maker of a database which is made available to the public in whatever manner may not prevent a lawful user of the database from extracting and/or re-utilizing insubstantial parts of its contents, evaluated qualitatively and/or quantitatively, for any purposes whatsoever.

Where the lawful user is authorized to extract and/or re-utilize only part of the database, this paragraph shall apply only to that part".

- The database must have been made available to the public "in whatever manner". This is considered to also include a database which is not freely accessible but only on subscription. The users will then only be lawful if they have granted subscription.
- This exception is considered to be broader that what falls under the notion of a "normal use" under Article 6.1. "Any purpose whatsoever" includes purposes which the maker may not have intended, including "non normal" uses.[278]

#### 5.8.4.4 Repeated and systematic extraction or re-utilization of insubstantial parts

While the use of insubstantial parts remains outside of the protection scheme, a special rule has been inserted in the Directive to prevent 'circumvention' of the protection scheme by users that would systematically retrieve insubstantial parts. According to Article 7(5), such behavior does amount to infringement, albeit only insofar the extracted parts constitute a substantial part when taken together.

[277] British Horseracing Board Ltd v William Hill Organization Ltd (BHB decision), C-203/02.
[278] See De Wolf & Partners, *Study on the legal framework of text and data mining*, 2014, p. 77.

### 5.8.5 The exception for scientific research

As is the case in copyright law, the *sui generis* protection scheme includes an – optional – exception for purposes of scientific research by lawful users.[279] This use is allowed on the traditional condition that the source is indicated and that the use is justified by the non-commercial purpose of the research.

Article 9(b) of the Database Directive: "Member States may stipulate that lawful users of a database which is made available to the public in whatever manner may, without the authorization of its maker, extract or re-utilize a substantial part of its contents:

[…] (b) In the case of extraction for the purposes of illustration for teaching or scientific research, as long as the source is indicated and to the extent justified by the non-commercial purpose to be achieved".

The research exception is an exception to the right of extraction and not to the right of reutilization. Recital 51 states that Member States may "permit a lawful user of a database to extract a substantial part of the contents for the purposes of illustration for teaching or scientific research,"

When data is extracted as a lawful user from a database the following criteria have to be met to qualify under the exemption namely:

(i) data are extracted for the purpose of illustration for teaching or scientific research;
(ii) the source is indicated;
(iii) Data are extracted to the extent justified by the non-commercial purpose to be achieved.

Article 9(b) of the Database Directive does not include the adjective "sole" purpose of scientific research so also other non-commercial purposes, can benefit from this exception, which is relevant for OpenScienceLink.[280]

Nevertheless, and considering the aims of the OpenScienceLink platform, the chances to rely on this exception, appear very limited. We refer in this respect to the concern that was expressed by the academic and scientific community because of the fact "that the exceptions to the "*sui generis*" right were too restrictive with regard to the access to and use of data and information for scientific and educational purposes".[281]

### 5.8.6 Conclusion

OpenScienceLink aims to develop a universal comprehensive and well-structured repository of scientific and research data. As regards the protection of the repository by the *sui generis* right, a positive answer will depend on the existence of 'a substantial investment' which seems likely, even though some legal uncertainty persists.

The repository will clearly be making use of substantial parts of other databases. Depending on whether or not the latter can benefit from the *sui generis* right, prior permission should be requested from the

[279] On the concept of 'lawful user', see *supra*, Section 5.7.3.1.
[280] See for detailed analysis De Wolf & Partners, *Study on the legal framework of text and data mining*, 2014, p 81.
[281] EC Working Paper, *First evaluation of Directive 96/9/EC on the legal protection of databases*, 12 December 2005, p. 21 (available at http://ec.europa.eu/internal_market/copyright/docs/databases/evaluation_report_en.pdf)

rights owner of the database. Even when OpenScienceLink will only be using non-substantial parts of third party databases, the repeated or systematic retrieval of such parts may be qualified as an infringement. Prior agreements seem advisable. Such agreements should preferably also include the possible 'copyright part' of the protection, provided that it is owned by the same entity.

We are not the first to observe that the complexity of the "*sui generis*" regime due to the two tier approach of the Directive has caused confusion among users as the same database can be protected by both copyright and "*sui generis*" right. In particular, the association of European academies represented by ALLEA ("All European Academies") revealed serious concerns about the effect of the Directive upon scientific research. [282]The main concern is that the Directive limits access and the use of data and information for scientific and educational purposes. This is held to impede research and reduce the public benefit which might otherwise be derived from research. In the view of ALLEA, the Directive is designed for the commercial sector whilst scientific data and the way in which scientists have traditionally used it is different in many ways.[283]

## 5.9 Specific use cases[284]

The most relevant types of usages based on the services provided by OpenScienceLink are access, linking, text and data mining, re-use and value added modifications and enhancements. These topics will be further addressed under work package 8 and will take into account the successive developments of the platform services. The text below already sketches some main conclusions.

### 5.9.1 Linking

One of the tools provided by OpenScienceLink is the publishing and sharing of publications and experimental datasets, as well as linking them with researchers and scholars. In particular peer reviewed literature and associated datasets and collections as well information about the researchers and relevant journals etc. will be mutually linked. It should thus be examined whether such act of linking to information without the consent of the respective right holders could constitute an infringement.

Various courts have confirmed that a hyperlink is not a relevant reproduction of a work, because it does not constitute a new physical fixation but merely an electronic referral. With respect to the right of communication to the public, the CJEU has held that a clickable link to works that are freely available on another website does not constitute as an act of communication to the public. [285].

However, in order to make this linking technically possible the researcher will need to be able to access datasets and technically place a link between the different sources. The OpenScienceLink system will likely use an external database to resolve the relevant DOI's. That database may be protected by a *sui generis* database right. Processing the data of the DOI implies that a copy of this data is made. Although the individual linking by a researcher will not constitute a substantial part, OpenScienceLink Platform

[282] EC Working Paper, *First evaluation of Directive 96/9/EC on the legal protection of databases*, 12 December 2005, p. 22.
[283] ALLEA's letters addressed to Internal Market Commissioners Bolkestein (2002) and McCreevy (2005).
[284] This part of the analysis is inspired by a similar analysis in the framework of the OpenAireplus infrastructure; see Lucie Guibault and Andreas Wiebe (Eds.) *Safe to be open, Study on the protection of research data and recommendations for access and usage*, OpenAIRE Project, Universitätsverlag Göttingen, 2013.
[285] We refer to the previous paragraph 5.4.3.2.

as repository will (let) perform such acts for all the researchers using this service. It is thus likely that such accomplishment amounts to a (non-authorized) repeated use of insubstantial parts. A license for extraction from external databases seems therefore appropriate.

### 5.9.2 Harvesting

One of the goals of the OpenScienceLink platform is to enable harvesting of scientific data and establish connections.

The Platform keeps track of papers that are referring to or have used datasets of the OpenScienceLink repository. It is able to do so by linking publications to these datasets. OpenScienceLink incorporates relevant datasets from other repositories such as PubMed. The exclusive rights of extraction and re-utilization of the right holder of the *sui generis* database right in the PubMed and other databases (see list in annex) could be infringed by the intended use.

If the goals is to permanently incorporate data of external databases into the OpenScienceLink repository, it should be kept in mind that the act of copying data out of protected databases and into another database constitutes an act of extraction in terms of the *sui generis* database right. If the data being copied is a substantial part - which will be decided on a case by case basis by the courts - unauthorized use would lead to an infringement. It is unlikely that OpenScienceLink can rely on the exception for scientific use because, first, it is only applicable if the relevant national legislation provides for it and (second) if the copying is carried out for own scientific use. Since the copying is being done by OpenScienceLink as a legal entity the scientific research exception may not be applicable.[286]

An important point of attention concerns the copyright that may be attached to certain data that are being copied out of external databases, such as full text papers and elaborate metadata. Incorporating them into the OpenScienceLink database is being done without permission of the respective copyright owners would clearly amount to infringement.

It is crucial for the success of the OpenScienceLink platform to have access to external research data content systems. In order to be allowed to harvest not only scientific data but any information to optimize the review system proposed OpenScienceLink needs to clear authorization. This may result from the conditions of the license agreement that is attached to the database or should be obtained by an agreement with the rights owner. Since it is one of the aims of the many Open Access policy repositories to disseminate their contents as widely as possible, an authorization will often result from the Creative Commons 4.0 license that is commonly applied by such OA repositories.

In general, it remains very important to carefully examine the applicable license conditions in order to ascertain which restrictions are imposed by the copyright holders to the free exchange of information. OpenScienceLink keeps an overview of all the databases it will be using with an indication of the applicable license condition.

[286] It should be noted, though, that the precise contours of the research exception remain unclear and would need to be clarified by the CJEU.

### 5.9.3 Text and Data Mining

In order to achieve a universal well-structured repository of scientific and research data for experimentation and benchmarking of pertinent research works, tools such as text mining will have to be used. In the case of text and data mining, tools are used to read and classify research data and to analyze and identify connections and correlations. The results are stored in databases and can be accessed by the end users.

First, in so far as data used for mining activities is *copyright* protected, some exclusive rights of the holders of copyright may have to be taken into account.[287] It will have to be further examined which kind of actions are undertaken by the mining tools. The mere act of reading does not constitute an infringement. A reproduction or extraction of data (e.g. in memory of a computer) will concern the exclusive right of reproduction.[288] Using the work by making changes to it, or to adapt it into another work may moreover make the user liable for infringement of the exclusive right of adaptation.[289] Data mining does, however, in general, not involve a modification of the original data which are only read, classified and sometime extracted. Such activity will therefore normally not constitute an act of adaption.

Second, as regards the possible infringement of the *sui generis right* by applying text and data mining tools to protected databases, such will depend on the nature of the tools that are used. The *sui generis* right is infringed where a substantial part of the database transferred to another medium or made available to the public. However,

*"Within the sphere of data mining activities, data is not made publicly available but just read, classified and potentially copied. Whether the extraction right is infringed or not depends on what actions are carried out by the text mining tools*"[290]

Both in the case of the copyright framework as in the case of the *sui generis* scheme, it remains unclear to what extent the scientific use exemption may apply to infrastructures such as OpenScienceLink that make use of automated text mining tools.[291] It is not advisable to rely on such an outcome. In many Member States, the scope of application of this exception is limited to physical persons for their own scientific use.[292]

Hence, it is advisable to seek a license for the application of text and data mining tools to third party databases.

---

[287] Lucie Guibault and Andreas Wiebe (Eds.) *Safe to be open, Study on the protection of research data and recommendations for access and usage*, OpenAir Project, Universitätsverlag Göttingen, 2013, p. 107.
[288] *Ibid.* p. 108.
[289] The right of adaptation as such is not explicitly regulated on EU level although this remains unclear. This right is however protected on the international level by Article 12 of the Berne Convention.
[290] Gauibault and Wiebe, *op.cit.* 2013, p. 109.
[291] Gauibault and Wiebe, *op.cit.* 2013, p. 108: "it is virtually impossible to introduce an e-infrastructure whose reproductions would completely fall within the scope of the exception for scientific use in every European country".
[292] *Ibid.*, p. 110.

### 5.9.4 Re-use in different contexts

The OpenScienceLink platform provides not only services for researchers to publish and re-use publications and datasets but also to review, complement and enrich data by the different stakeholders using the tools within OpenScienceLink. Tools for the re-use of data include the analyses of data for trend reports but also combining different datasets for new research publications.

The following exclusive rights apply to re-use.

(1) If the re-use includes creating a new physical fixation of the copyright protected work the right of *reproduction* prevents this without authorization.

(2) As long as the work is used as is and not translated or rewritten there will be no illegal act of adaptation. The right of adaptation may, nevertheless, be relevant because within the OpenScienceLink structure the contained data will be reviewed, enhanced and enriched to create a well-developed repository. The aim of the OpenScienceLink repository is not to modify the data as such, but to create a repository where the data can be stored and accessed by its stakeholders who in turn can modify the data.

(3) The platform allows free access to all online users and as such makes a *public communication of the work*; the act of mere linking existing datasets, publications and associated datasets or sets of raw data will however not constitute a prohibited act as long as these data are not duplicated.

(4) The re-use, value-added modification, trend analysis and enhancements of protected data within a database does not infringe the *sui generis* database right because adaptation as a right is not mentioned as such.[293] The preceding act of copying and making available of a substantial number of such data however is. The question whether the re-use was 'substantial' in nature, fact-based and has to be answered on a case-by-case bases.

### 5.9.5 Conclusions

It is clear that the legal status of 'research data', datasets and databases is complex and uncertain under the scope of European law. This uncertainty could be alleviated, to a certain extent, through the use of contractual arrangements. This will be discussed in the next chapter

[293] Gauibault and Wiebe, *op.cit.* 2013, p. 112.

# 6 Contractual agreements

## 6.1 Licensing framework

The conclusions from the previous section, have made clear that the legal status of protection of research data publications, datasets and databasis is complex. As a consequence, caution is recommended for the use of research data within the OpenScienceLink project. A clear licensing framework which takes into account the different types of research data and its forms of protection can help to avoid liability for infringing acts as well as to inform the users about the possibilities to re-use the research data.

OpenScienceLink needs to identify which research data is protected and may require rights clearance in the form of a license. [294] For publications and datasets which are not protected, there will be no ground for licensing. [295]

A range of standard licences are available, which can be used to achieve the goal of OpenScienceLink, i.e. open data. The use of an existing licence has many benefits, including enhanced organisational efficiency, cost saving and greater interoperability of data as well as increased user awareness of the licence terms, thereby enabling better compliance. OpenScienceLink will choose a license for the platform as a whole and a specific licence for the various types of research data that will become the content of the platform.

### 6.1.1 Open Data Licensing

The goal of OpenScienceLink is to publish and make available the data and datasets under conditions of Open Access.

Open Access entails "the free availability of scientific literature on the public internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. The only constraint on reproduction and distribution, should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited."[296]

The rationale given by many policy makers and scholars for pushing open access conditions is that “opening up the data allows for new knowledge to be discovered through comparative studies, data mining and so on. It allows greater scrutiny of how research conclusions have been reached, potentially

[294] Ball, A. (2012), How to licence research data, *JISC legal*, (4) 16
[295]Gideon Emcee, C. (2009), Building a sustainable framework for open access through information and communication technologies, *IDRC-CRDI*, (29), p. 41.
[296] Budapest Open Access Initiative (2002). Retrieved at http://www.budapestopenaccessinitiative.org/read

driving up research quality."[297] This rationale is also the idea behind the OpenScienceLink project; "to foster the widest access and re-use of scientific publications and data".

To support this open access business model, open data contracts and open access licenses were developed. These models increase legal certainty for potential re-users as to what extent they are authorized to use data that can be found online. Merely uploading research data on the Internet (even in the absence of any condition restricting use) does not in any way derogate from or call into question the author's exclusive rights. It moreover does not provide any information as to the author's intention to make the data available for open access. Without a license or terms of use, potential re-users are left with confusion and uncertainty.[298]

An Open Content license is a contractual agreement under which data, protected by intellectual proprietary rights, are made available openly for use or re-use subject to terms and conditions specified by the rights holder. Apart from the development of bespoke licences to facilitate the use and potential reuse of data, there is also a range of standard licences, which can be used to help achieve open data.

It is often more beneficial to use standard licences rather than bespoke ones. Apart from the benefits of enhanced organisational efficiency and cost saving, the use of standard licensing terms can lead to greater interoperability of data as well as increased user awareness of the licence terms, thereby enabling better compliance[299]. The most recognised and used standard licences, which can be used for data and datasets are summarised below.[300]

### 6.1.2 Creative Commons version 4.0

Creative Commons licences are standardized, automated licenses that authors can affix to their work in order to indicate under which conditions it may be used[301]. Creative Commons licences are originally intended for most copyright protected works (music, photographs, videos, etc.) and intended for every kind of author (musicians, photographers, filmmakers, writers, etc.)

The licenses are widely used for the dissemination of scientific publications under open access conditions. The current version of Creative Commons is the version 4.0 which now also addresses the *sui generis* database right.

---

[297] Ball, A. (2012) "How to License Research Data", *DCC How-to Guides*, Edinburgh: Digital Curation Centre. Available from http://www.dcc.ac.uk/resources/how-guides; see also Communication from the Commission of 17 July 2012, *Towards better access to scientific information. Boosting the benefits of public investments in research*, COM (2012) 401 final (1).

[298] Gideon Emcee, C., *op.cit.* 2009, p. 26.

[299] Note that not all of the licenses are conformant with the principles laid out in the so-called Open Definition (http://opendefinition.org/od/). Licenses conform to the Open Definition if they are: Reusable: Not specific to an organization or jurisdiction. Compatible: Must be compatible with at least one of GPL-3.0+, CC-BY-SA-4.0, and ODbL-1.0. Permissive/attribution-only licenses must be compatible with all 3 of the aforementioned licenses, and at least one of Apache-2.0, CC-BY-4.0, and ODC-BY-1.0. Current: Widely used and generally considered best practice by a broad spectrum of projects and actors within the domains of applicability of the license.

[300] See also a discussion thereof in Korn, N. & Oppenheim C. (2011), *Licensing Open Data: A Practical Guide*, available at http://discovery.ac.uk/files/pdf/Licensing_Open_Data_A_Practical_Guide.pdf

[301] Available at www.creativecommons.org

The main CC licences offer a series of 'baseline rights', with attribution (CC BY) as a core requirement, together with three other 'licence elements' that can be mixed and matched to produce a customised licence through a point-and-click web interface.

In the order of least to most restrictive licence these are the following possible combinations of CC licences:

| | | | |
|---|---|---|---|
|  | CC BY | **Attribution** | CC licence which allows third parties to copy, distribute, display, and perform the work as long as the creator is given due credit. The licensor has to specify the way in which credit is given. |
|  | CC BY-SA | **Attribution Share-Alike** | The SA clause inserts a strong copy left clause into the licence. All derivative works must be released under the same licence as the original work. |
|  | CC BY-ND | **Attribution No Derivatives** | The licensee is prohibited from altering, transforming or building upon the work. When the ND clause applies, the Share Alike clause cannot apply. The two are mutually exclusive. |
|  | CC BY-NC | **Attribution Non-Commercial** | The licensee cannot use the work for commercial purposes. Derivative works do not have to be released under an identical non-commercial license. |
|  | CC BY-NC-SA | **Attribution Non-Commercial Share-Alike** | Licensee cannot use the work for commercial purposes. All derivative works must be released under the same license as the original work. |
|  | CC BY-ND | **Attribution Non Commercial No derivatives** | The licensee is prohibited from altering, transforming or building upon the work. Sharing the work is authorized but only for non-commercial purposes. |

**Table 1:** www.creativecommons.org

The copyright holder is free to attach any of these conditions or a compatible combination of them to the license.


CC0 is the 'public domain' licence enabling scientists and other owners of copyright- or database-protected content to waive all their intellectual property rights and thereby place them as completely as possible in the public domain. However since the legality of this license is debated, inter alia because of the principle of inalienability of moral rights, the CC0 licence should not be chosen for the OpenScienceLink Project.

### 6.1.3 Open Data Commons

A more recent project in the realm of open access is the Open Data Commons Project, set up in 2007 by the Open Knowledge Foundation.[302] These licences apply to more radical forms of open data, whereby rightholders are required to dedicate their works to the public domain for the benefit of the public, and relinquish all rights in the work.[303]

[302] Retrieved from http://opendatacommons.org
[303] Gideon Emcee, *op.cit.* 2009, p. 31.

The Open Data Commons Attribution licence (ODC-BY) and the Open Data Commons Open Database Licence (ODC-ODbL) are specifically geared towards databases and their contents protected by copyright or *sui generis* database rights

These are the Open Data Commons licences currently available, leaving out the Public Domain Dedication and License.

| | | |
|---|---|---|
| ODC-By | **Open Data Commons Attribution Licence** | - Allow licensees to copy, distribute and use the database, to produce works from it and to modify, transform and build upon it for any purpose.<br>- You must attribute any public use of the database, or works produced from the database, in the manner specified in the license. For any use or redistribution of the database, or works produced from it, you must make clear to others the license of the database and keep intact any notices on the original database.<br>- Worldwide, royalty-free, non-exclusive licence to use the database for the duration of any applicable copyright/ database rights.<br>- Non-commercial and commercial exploitation. |
| ODC-ODbL | **Open Data Commons Open Database Licence** | - Allows use cf. ODC-By<br>- Attribution requirement<br>- World-wide, royalty-free, non-exclusive<br>- Non-commercial and commercial exploitation<br>- Share-Alike requirement: if you publicly use any adapted version of this database, or works produced from an adapted database, you must also offer that adapted database under the ODbL<br>- Openness requirement: If you redistribute the database, or an adapted version of it, then you may use technological measures that restrict the work (such as DRM) as long as you also redistribute a version without such measures |
| ODC-DbCL | **Open Data Commons Database Contents Licence** | - The Open Data Commons licences in general only cover the *database* itself, and not it's content. Therefore, it is not possible to only use one licence for the whole database including its contents. The ODC-DbCL can be used in conjunction with the ODbL to waive copyright for the contents of the database.<br>- The contents licence is formulated in an extremely broad manner, making this licence 'empty' in many jurisdictions.<br>- Attribution requirement<br>- Share-Alike requirement<br>- Openness requirement |

**Table 2:** www.opendatacommons.org

Given the fact that the purpose of OpenScienceLink is to publish scientific publications together with datasets upon which these publications are based, it would be necessary to use two different licences (or more), one for the database and one for its contents. This is unpractical and will likely create confusion and incompatibilities.[304]

### 6.1.4 Digital Peer Publishing License

The Digital Peer Publishing Licence (DPPL) is a licence developed in Germany especially for text works or electronic journals. It was created in 2003 on the initiative of the State of North Rhine-Westphalia to promote the foundation and spread of e-journals. It is specifically designed for scholarly content because

[304] Guibault, L. (2012), (19)

it covers all aspects of authenticity, citation, bibliographic data and metadata, permanent access and open formats.[305]

The licence comprises of three core modules: the basic licence model (DPPL, akin to the CC-By-ND licence), the modular licence and the free licence (akin to the CC-By licence). All three licences guarantee proper attribution of the author. They are nearly identical with the exception of the permissibility of (1) the physical distribution of the work, especially in print media, and (2) modifications to the work. While the basic DPPL and the modular DPPL allow only the distribution of works in electronic form and permit users to make them available for download, the free DPPL also allows works to be distributed in physical form, especially in print media.[306]

The basic Digital Peer Publishing Licence module allows users to read the work, to electronically distribute verbatim copies and to make it available for download. The DPPL restricts the possibility of the user to make modifications. This is what makes the basic DPPL particularly suitable for making available to the public finished works which the author wishes to be distributed verbatim with due attribution. It is what also makes this licence the most restrictive licence compared to the other two.

The basic module makes no distinction between scholarly or commercial use of works. The work can only be distributed in electronic form, so the rights of use in print form or on data storage media are retained by the author and not covered by the licence. The idea behind this licence is thus to promote electronic distribution, leaving the right holder the option to grant third parties the right of physical distribution.

The second licence, the Modular Digital Peer Publishing Licence, is less restrictive than the basic licence. It enables the author to allow modifications to the work, but only in the areas designated by the right holder as suitable for modification. The other parts, not specifically earmarked, can then only be distributed verbatim. This way, individual graphs of formulae might be changed, while the text remains unchanged.

The third licence, the Free Digital Peer Publishing Licence, is the least restrictive licence of the set. It allows work to be altered and permits the distribution of modified versions. Hence, this licence is suitable for situations where several scholars collaborate on a work. Because of the attribution requirement, reference should always specifically be made to the original author, even for modified versions unless the author decrees that his or her name should not be mentioned. Moreover, modified versions must also be distributed under a free DPPL (supra: openness requirement) thereby ensuring that the author of the modified version does not benefit one-sidedly from the generosity of the original author. The idea behind these licences is to encourage open access and bring about new scientific developments.

These licences are very suitable for copyright-protected works. However, it is doubtful whether these licences are adapted to scientific data or datasets. Databases are not mentioned in the scope of these licences and they are specifically geared towards copyrighted works.[307]

[305] Retrieved from http://www.dipp.nrw.de

[306] *Ibid.*

[307] *Ibid.*

## 6.2 Recommendation

### 6.2.1 Challenge

Open Access encompasses a range of components such as readership, reuse, copyright, posting, and machine readability. [308]

If we look at the requirements relating to open access, data protection and intellectual property law, a balance needs to be found between, the one hand, a restrictive license to safeguard privacy and avoid infringements and, on the other hand, the aim of open access making works freely available without restrictions.[309]

OpenScienceLink needs to comply with third party rights when it wants to make data available under the chosen license. As use conditions and restrictions may change over time, there will be a constant need to re-evaluate the chosen licence. This may also be necessary in the light of further developments and new goals of the OpenScienceLink platform services. For the time being, licence compatibility and interoperability remains a difficult issue. Moreover, as the license chosen for the OpenScienceLink project does not constitute the most open license, changes may be needed in the future taking into account further developments of the project. [310]

### 6.2.2 Choice for Creative commons CC04 AT-NC-SA

With the launch of Creative Commons 4.0, this type of license appears to be most suitable for OpenScienceLink. As a globally recognizable license, it brings not only instant recognition and assurance of the legal status but it also assures some interoperability of the content. Researchers in the EU are assured that their rights and permissions will be recognized more easily by users outside the EU and vice-versa.

Adding to that the inclusion of the *sui generis* database right in the CC 4.0 version, its interoperability has only increased. Therefore it constitutes a suitable license under which the research data included in the OpenScienceLink platform will be made available.

---

[308] See for an overview http://sparc.arl.org/sites/default/files/hoii_guide_rev4_web.pdf

[309] Open Access Clauses in Publishers' Licenses: Current State and Lessons Learned https://www.coar-repositories.org/files/OA-Clauses-in-Publishers-Licenses.pdf

[310] Cf. the Elsevier Text and Data Mining (TDM) License. One of the reasons to only allow for non-commercial use has to do with Academic subscribers: Researchers can text mine subscribed content on ScienceDirect for non-commercial purposes, via the ScienceDirect API's. Text and data mining enabling clauses for non-commercial purposes will be included in all new ScienceDirect subscription agreements and upon renewal for existing customers. Librarians interested in adding the TDM clause to their existing agreement prior to renewal are able to request a simple contract e-amendment via their Elsevier Account Manager.

**Licensing issues with derived data**

In the context of the OpenScienceLink project, different types of data will be included. There are vast quantities of cross-disciplinary base data, which are generated by automatic tools and human intervention.

The goal of OpenScienceLink is to assemble and organize this data with the open access purpose in mind. However, it is also possible that tools such as text mining, data mining and other analysing operations will be used or that this data will be enriched with other data, before it is made available on the OpenScienceLink platform. The output from these operations is generally referred to as 'derived data'.
These activities raise complex intellectual property and licensing issues, including the potential creation of new intellectual property rights from multiple texts/datasets. An example has been given earlier in this deliverable, where we suggest that the creation of the OpenScienceLink database with collected base data/derived data might generate *sui generis* database rights.
This section will present a brief overview of most IPR and licensing issues, which have to be taken into account when producing derived data[311]:

(a) (1) At the base level, three intellectual property rights can and will probably exist in the original documents:
    (a) Copyright in the original text;
    (b) Database rights for those platforms which have invested substantially in the collection and arrangement of the base data (e.g. GeoPubMed and PubMed); and
    (c) Moral rights (these rights protect *inter alia* the reputation of the author)

↓

(b) (2) To obtain this base data, OpenScienceLink will likely need consent from all the owners of the rights to the works and/or collections to assemble and analyse the original documents/base datasets.

↓

(c) (3) In the course of collecting and assembling these documents/data, it is possible that the substantial effort of OpenScienceLink actors will generate *sui generis* database rights on this collection.

↓

(d) (4) When applying the analysing operations, or adding other base data, a new 'derived' data set will have been created.

Identifiable legal issues:
    (a) Check licensing framework for each of the original datasets: what restrictions are there on the re-use and processing of this data?
    (b) Attribution to the copyright owner of a work or collection
    (c) Who owns the rights in the resulting dataset? The employer; research institution, university or enterprise, or the employee? Is this a joint ownership with owners of rights in the original datasets?
    (d) What licensing framework do we use on the 'derived' dataset?

↓

(e) (5) The user submitting queries through the datasets will use the results of this query on the platform. The user will either be the owner of the derived dataset or the user will have agreed to
    (a) Use the derived dataset under specific licensing conditions and
    (b) Use the platform in accordance with the terms and conditions of the platform.

[311] For more detail on this roadmap: Korn, N., Oppenheim, C. and Duncan, C. (2007), IPR and Licensing issues in Derived Data. Retrieved at http://www.jisc.ac.uk/media/documents/projects/iprinderiveddatareport.pdf

# 7 Liability of Internet Service Providers

## 7.1 Overview

The role of Internet service providers (ISPs), is to 'provide access to, host, transmit and index content originated by third parties on the Internet; facilitate interactions or transactions between third parties on the Internet; or provide other Internet-based services to third parties.

Because these services of intermediaries may increasingly be used by third parties for infringing activities ISP's are also considered to be best placed to bring such infringing activities to an end. Therefore, right holders whose rights are being infringed should have the possibility of applying for an injunction against an intermediary.[312]

The relevant European directive in this respect is the Directive of 8 June 2000 on E-commerce (OCD).[313] It regulates several aspects of information society services, including freedom of services, the treatment of electronic contracts, and liability issues for third party content, among others. In this section we briefly present the scope of the Directive before focusing more on the intermediary liability provisions that may be relevant for the OpenScienceLink platform.

A distinction is made between primary liability which governs the act of infringement of the ISP itself and liability which covers secondary infringements, i.e. liability that is placed on a person who did not directly commit the infringement but who helped the infringer or benefited from the infringer. For some cases of secondary liability, the e-commerce Directive offers a 'safe-harbour', exempting certain ISPs from liability. It is thus important to investigate if and to what extent OpenScienceLink can be exempt from liability for infringing use by third parties that occurs through its platform.[314]

It should be noted that, even in the case of an exemption from liability, rights holders are in a position to apply for an *injunction* against intermediaries whose services are used by a third party to infringe an IPR.[315]

[312] Zittrain J. (2006), A History of Online Gatekeeping, Harvard journal of Law and Technology, Vol. 19, No. 2, 258; Wong, C. and Dempsey, J.X (2011), Mapping Digital Media: The Media and Liability for Content on the Internet, *Open Society Foundation*, Reference Series No.12, 14

[313] Directive 2000/31/EC of the European Parliament of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market ('ECD'), OJ L 178,, 1-16

[314] We will not discuss the implementation of the E-directive within member state legislation, so certain issues fall outside of the scope of this chapter.

[315] See Art. 11 Enforcement Directive and Art. 8 (3) InfoSoc Directive.

The provisions of the Directive have been almost literally transposed into national law by the Member States.[316] Some Member States have expanded the waiver of liability to include hyperlinks and search engines.[317]

## 7.2 The internet service provider (ISP)

To be able to decide whether OpenScienceLink can benefit from the liability exceptions it should be considered an 'internet service provider' in the meaning of the ECD.

The ECD uses the term ISP to describe any operator who provides a wide range of online services. Art 2(b) defines an intermediary 'service provider' as any natural or legal person providing an information society service.

Information Society Services are defined as *'any service normally provided for remuneration, at a distance, by electronic means and at the individual request of a recipient of services'*. [318]

The key elements to determine if a service qualifies under the directive are

**Remuneration.**

> Information society services span a wide range of economic activities, which take place online.[319] A service which is subject to payment for example the paid service of Google AdWords, falls under the scope of the requirement. Recital 18 however explains that payment is not required as long as the service "represent an economic activity". Access fees or associated advertising are further examples[320]
> When a service is free and without advertising, this may still meet the requirements. Also non-commercial services such as those provided by governmental agencies and NGO's are considered economic activities. As long as the service is broadly part of an economic activity, the criteria is fulfilled. [321]

**At a distance;**

---

[316] Report from the Commission to the European Parliament, the Council and the European Economic and Social Committee on the application of Directive 2000/31/EC of the European Parliament and of the Council on 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market (Directive on electronic commerce), COM/2003/702, 12-13.

[317] Ibid, 13.

[318] Article 2 of the E-Commerce Directive.

[319] These activities consist in particular of selling goods on-line; transmission of information via a communication network, providing access to a communication network, hosting information provided by the recipient of the service, services providing search tools, on-line contracting, etc. Recital 18 of the e-Commerce Directive

[320] Lodder A. (2002), 'Directive 2000/31/EC on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market, in Lodder A. and Kasspersen (eds.), *eDirectives: Guide to European Union Law on E-commerce - Article by Article comments*, Kluwer Law International, 71.

[321] *The element of remuneration is very broadly interpreted. At this point it refers to the existence of any economic activity or an activity for which an economic consideration is given in return. Information society services therefore extend to services which are not remunerated by those who receive them. It also includes advertising and sponsoring which is often the case with social networking sites or a search engine, which are thus also information society services.*

To serve at a distance means that the service provider and user cannot be physically present.

**By electronic means;**

This is generally considered the case when both the sender and receiver use data processing equipment and where the service is transmitted via an electronic communication device such as a computer or smartphone.

**At the individual request.**

The service must be provided to the recipient, such as the users of the platform, by individual request. Recital 18 refers to on demand services when the user has control over the information which is transmitted. Also visiting a website is considered to fall under the criteria, since the visitor requests the website through typing in the URL or clicking on a link.[322]

**Of a recipient**

The user as a recipient as defined in art 2(d) is 'any natural or legal person who, for professional ends or otherwise, uses an information society service, in particular for the purposes of seeking information or making it accessible;"
Recital 20 explains that 'The definition of "recipient of a service" covers all types of usage of information society services, both by persons who provide information on open networks such as the Internet and by persons who seek information on the Internet for private or professional reasons.' It is possible that an ISP is both a provider of a services as well as a recipient making use of a service, for example relying on an internet access service to make its own services available.

The E-Commerce Directive excludes a number of services and legal issues from its scope which are for example covered by the Data Protection Directive [323]

OpenScienceLink although its services are not paid for, provides user services that may qualify as internet society services. These include storage, hosting, and sharing research data, the possibility to use the service for peer review open access journals and to search and use trends within the researchers' field of interest. These services can also be used to share illegal information or unauthorized storage. For example, the peer review services may be used to spread defamatory statements and the search reports may show unauthorized third part content by mistake.

Given its broad definition, many of the economic activities that take place online fall under the scope of the E-Commerce Directive. Services developed with respect to OpenScienceLink include providing access to a communication network, hosting of information and offering on-line information and providing search, access and the use of retrieval of data tools. When OpenScienceLink qualifies as an ISP for these services then it may also benefit from the liability exception for these services. It is important to note that if for some services the conditions are not met, this does not mean that the

[322] Lodder A. (2002),

[323] Additionally, the Directive does not apply to: issues related to taxation; questions relating to agreements or practices governed by cartel law; the activities of notaries or equivalent professions to the extent that they involve a direct and specific connection with the exercise of public authority. See article 5.1 E-Commerce Directive.

intermediary is always subject to liability. It only means that the intermediary cannot rely on the immunity provided by the Directive. In that case the question of liability will be determined under the applicable material law specific for the type of infringing content in each Member State.[324]

## 7.3 Obligations of the Information Society Service Provider

According to Article 5 of the ECD, any Internet service intermediary has the obligation to provide certain information towards the users of the information society service and the competent national authorities.[325]

## 7.4 Liability Exemptions for Intermediaries

Section 4 of the ECD regulates the liability framework of certain services of intermediary service providers (ISP). The ISPs will not be held liable for three specific services: 'mere conduit' (article 12); 'caching' (article 13) or 'hosting' (article 14). In order to benefit from these exemptions, the providers must comply with the conditions listed in each article.

The rationale behind these exemptions for intermediaries was the concern that ISPs should not be held liable for third party content on similar grounds as 'publishers'. These 'liability havens' were set up in order to encourage service providers to enter the market.[326]

It is important to note the distinction between the service provider and its activities. The immunity is given to particular services and not to the ISP itself.[327]

### 7.4.1 Mere conduit

The first exemption of liability concerns the situation of 'mere conduit'.[328]

This exemption includes two types of services:

- 'transmission services'; Those services which consist of the transmission in a communication network of information provided by a recipient of the service and

[324] Van Eecke P., Truyens M., *Legal analysis of a Single Market for the Information Society, New rules for a new age?* A study commissioned by the European Commission's Information Society and Media Directorate-General, November 2009. Chapter 6: Liability of Online Intermediaries, p. 10. Available at: *http://ec.europa.eu/information_society/newsroom/cf/document.cfm?doc_id=842*.

[325] Such as the name, address at which the service provider is established, and other details of the service provider, including his electronic mail address.

[326] Walden I. (2006), in Bullesbach A, Poullet Y., Prins C. (eds.), Concise European IT Law, Kluwer Law. International, p. 248.

[327] See First Report on the application of Directive 2000/31/EC [64] The Directive does not affect the liability of the person who is at the source of the content nor does it affect the liability of intermediaries in cases which are not covered by the limitations defined in the Directive. Furthermore, the Directive does not affect the possibility of a national court or administrative authority to require a service provider to terminate or prevent an infringement. These questions are subject to the national law of the Member States.

[328] Article 12 of the E-Commerce Directive.

- 'access services' Those services which consist of the provision of access to a communication network

In these two cases, the service provider will not be held liable for that transmission. Such is however, subject to the conditions that the service provider

- Does not initiate the transmission;
- Does not select the recipient of the data; and
- Does not select or modify the transmitted data.

Recital 42 further stipulates that the exemptions provided by the Directive apply only to cases 'where the activity of the information society service provider is limited to the technical process of operating and giving access to a communication network'.[329] It further elaborates that such activities are of mere technical, automatic and passive nature, which implies that the information society provider has neither knowledge of nor control over the information, which is transmitted or stored.

Even if such information transmission or access provision would include automatic, intermediate and transient storage, such storage would still be regarded as mere conduit if it is conducted for the sole purpose of transmitting information and if the information is not stored for a longer period than necessary for the transmission.[330]

Although the ISP cannot be held liable for mere conduit, Member States may have national regulations allowing court or administrative authority to order the service provider to terminate or take measures to prevent a particular infringement.[331]

Given the restrictive nature of the conditions, especially the technical and passive nature of the activities, it is unlikely that OpenScienceLink platform would qualify for this exemption under the Directive. OpenScienceLink plays an active role in selecting repositories and validation of datasets so it will be difficult to argue that OpenScienceLink does not select or modify the information or does not have any control over the information that is transmitted and/or accessed through the platform.

### 7.4.2 Caching

The second liability exemption concerns 'caching' of information. Service providers who *automatic, intermediate and temporary store information* with the sole purpose of making the further transmission of the information more efficient are not held liable. This exemption does not extend to access services, unlike the 'mere conduit' exemption.

To qualify for this exemption the information may not be modified by the provider. Further requirements are that the provider

- Complies with conditions on access to the information;
- Complies with industry standard rules regarding updating the information,
- Does not interfere with the lawful use of technology, widely recognized and used by industry, to obtain data on the use of the information (hits); and

[329] Recital 42 of the E-Commerce Directive.
[330] Article 12 (2) of the E-Commerce Directive.
[331] Article 12 of the E-Commerce Directive gives the Member States this option.

- Acts expeditiously to remove or to disable access to the information it has stored, when he has actual knowledge of the fact that the information at the initial source of the transmission has been removed from the network, or access to it has been disabled, or that a court or an administrative authority has ordered such removal or disablement.[332].

It is not to be excluded that the OpenScienceLink platform will engage in such technical caching operations for which it could invoke the exemption. This should be further investigated.

### 7.4.3 Hosting

The third exemption relates to activities of 'hosting'.[333] An example of a hosting service is the popular video service YouTube.

If the service consists the storage of information which is provided by a recipient of the service, the service provider shall not be liable for the information stored on the condition that:

- The provider does not have actual knowledge of illegal activity or information, and as regards claims for damages, is not aware of facts or circumstances from which the illegal activity or information is apparent; or
- The provider, upon obtaining such knowledge or awareness, acts expeditiously to remove or to disable access to the information.[334]

Furthermore, the service provider may not control or exercise authority over the recipient of the service and he may still be required to remove the information or block access thereto.[335]

Not all hosting service providers will benefit from this safe harbour. The CJEU has limited the scope of application to ISP's "where that service provider *has not played an active role* of such a kind as to give it knowledge of, or control over, the data stored". [336]/[337]

It has to be seen whether this exemption could apply to OpenScienceLink  In case users are free to upload any content on the platform, without any moderation, selection of content producers or other active intervention then this exemption would apply. However, this seems unlikely as OpenScienceLink takes an active part in the selection of some of its content producers such as GeoPubMed for instance.

---

[332] Article 13 (1) of the E-Commerce Directive.
[333] Article 14 of the E-Commerce Directive.
[334] *Ibid.*
[335] Article 14 (2) and (3) of the E-Commerce Directive.
[336] CJEU, 23 March 2010, Google/Vuitton, para 120.
[337] See, e.g., a decision by the French Supreme Court (Fr. Cass.) of 3 May 2012 that refused to exempt eBay from liability under Art. 14 ECD. An opposite decision was reached as regards the services of YouTube by Landgericht Hamburg, 20 April 2012, Gema/YouTube. See also, Madrid Court of Appeal, decision No 11/2014, case YouTube v Telecinco. In this case Telecinco alleged that YouTube operated as a content provider with an editorial control over the website. However the court did not agree that the facts in the case implied that YouTube was acting in a non-passive way, in the sense required by the Court of Justice to exclude the applicability of the liability exemptions.

### 7.4.4 No general obligation to monitor

With respect to information being transmitted or stored using a service covered by articles 12, 13, and 14 (i.e. mere conduit, caching or hosting) OpenScienceLink does not have a general obligation to monitor the legality of the content. Furthermore there is no obligation to actively seek facts or circumstances indicating illegal activity. Both these obligations would go against the idea that the ISP is neutral and passive with respect to the information and therefore receives limited liability for illegal content.[338]

Member States may however compel the ISPs to promptly inform the public authorities about illegal data or infringements reported by recipients of their services. They can also oblige the ISPs to communicate information enabling the disclosure of the identity of their subscribers with whom they have storage agreements at the request of public authorities.[339]

Finally, it is important to recall that, despite the lack of liability of the service providers, a court or administrative authority may still order the service provider to terminate or take measures to prevent a particular infringement where the legal system of the Member State provides for this under article 12 of the Directive.

[338] Article 12 and 15 of the E-Commerce Directive 2000/31 See also CJEU, Case C- 70/10, Scarlet/Sabam.
[339] Article 15 (2) of the E-Commerce Directive.

# 8 Conclusion

All the legal requirements presented in this deliverable have to be taken into account by the OpenScienceLink project. Fulfilling these requirements is necessary and obligatory in order to ensure the legal compliance of the platform and the different pilots with the regulatory framework. These requirements will be consolidated and completed here.

### 8.1.1 BASIC DATA PROTECTION REQUIREMENTS

Data Protection compliance

- The OpenScienceLink platform has been analyzed in light of the European Data Protection Directive requirements. The OpenScienceLink platform implemented all requirements relevant at this stage. Given the developments of the platform services and the continuing discussion on the Data Protection reform package, WP8 will provide further details incorporating the latest technical and legal developments. One of the aspects that will need further investigation for example concerns the processing of research data potentially containing personal data. At the moment the platform only allows for processing of non-personal research data. If however after review open access can be given to research data which holds personal data an access system with restricted access will have to put in place.

**First stage implementations include:**

Allocation of roles and responsibilities

- OpenScienceLink determines the purpose and means of the processing the personal data with regard to the user registration and user profiles kept and fed by the platform.
- Consequently OpenScienceLink is primarily responsible and liable for the legality of this part of the data processing. OpenScienceLink needs to respect data subjects' rights and need to appoint a contact person questions, request and complaints can be addressed to. As controller OpenScienceLink is responsible for the obligations towards the National Data Protection Authority (NDPA) and the data subjects.

Ensure legitimacy of processing

- The legal basis for processing of personal information of the platform users is the user's informed consent. OpenScienceLink will use an electronic consent form presented to the user before registering a user profile.
- The requirements for consent have been taken into account when developing the profile registration form.
- A versioning and archiving system must be in place for the informed consents given by data subjects to enable later verification that appropriate notice was given.

Ensure respect for data quality principles

- The OpenScienceLink platform implemented the general requirements on fair data processing, purpose-specification and proportionality, data minimization and information quality.

- When asked for user information it will be clearly indicated whether information is optional or required.
- In its privacy notice and terms of use OpenScienceLink Platform states the purpose of the data processing, taking into consideration if the data being processed is both adequate and non-excessive to achieve the purposes of the OpenScienceLink platform.
- Further use will be limited to articulated purposes.
- OpenScienceLink will limit storage duration whenever possible while also ensure that the user data is readily available to authorized entities as long as it is necessary.
- Procedures will be put in place to be able to delete information securely when no longer necessary for the purpose on how to report and deal with suspected inaccuracies. OpenScienceLink in its function as platform will rely on the data controller to delete or render the information anonymous when no longer necessary for the purpose and keep the information accurate and up to date. For the user data for which OpenScienceLink is the controller it will do the same.

Transparency and administrative requirements

- Through the use of website notices users of the platform services and data subjects are given an appropriate amount of information sufficiently detailed, but also comprehensible for the consumer.
- Transparency is provided on the rights of the data subject such as the right to rectify.

Security and confidentiality requirements

- The OpenScienceLink infrastructure should implement advanced organizational and technical security measures to ensure confidentiality, integrity and authenticity.

The OpenScienceLink platform in its current constellation falls under German law. The analysis of the implementation of European requirements in German law was not subject of this study. When further rolling out the platform compliance with German law will need further investigation.

### 8.1.2 COPYRIGHT and SUI GENERIS DATABASE RIGHT

- Whether the structure of the OpenScienceLink database will be held original and could benefit from copyright protection remains uncertain. Generally scientific databases lack the required level of originality but this is still a matter for the courts to decide on a case to case basis.
- The same question will arise as regards the possibility of *sui generis* database protection for the OpenScienceLink database. However, a positive reply seems more likely considering the substantial investment that is made in the setting up of the database. A further assessment of this issue will be made in work package 8.
- Use of publicly available databases, research data and repositories must be in conformity with their underlying licenses. Such licenses must in particular be examined with respect to the possibility to apply text and datamining; alternatively a new license agreement may need to be negotiated
- The intellectual property rights that pertain to scientific publications might be owned by different entities: the initial creator, the employer, co-authors or publishers. These rights might have been

further transferred to third parties as well, i.e. to the OpenScienceLink platform. This will need to be assessed on a case-by-case basis prior to incorporating any of these works into the OpenScienceLink platform. Moreover, clear guidelines in this respect need to be given to the users of the platform who wish to upload their data. Such guidelines will be developed under work package 8.

### 8.1.3 LICENSING FRAMEWORK

The licensing framework needs to reflect the complex demands of the OpenScienceLink platform and its potential users. It is therefore unlikely to propose a 'one license fits all' solution.

- Licensing conditions must conform to the requirements for open access, copyright and data protection. They must, moreover, be suitable for different data types including databases, be user friendly, sustainable and interoperable.
- Based on different use-case scenarios and services such as data mining, linking and review integrated into the OpenScienceLink platform the following license framework has been chosen as a starting point for development of the platform services:
    - The OpenScienceLink platform and its content (including uploaded research data) will apply the Creative Commons 04 license with the following restrictions: CC-BY-NC-SA.
    - Data extracted into the OpenScienceLink platform from external database repositories will apply the Creative Commons 04 license with the following restrictions: CC-BY-NC-SA, unless this would not be allowed by the underlying license.
    - As a condition to incorporate datasets, databases and copyright protected works, authors or other rights owners should explicitly agree to grant OpenScienceLink a license for open access use of their materials as envisaged by the platform services. As a minimum the CC-BY-NC-SA is required.[340]
- OpenScienceLink will only make use of publicly available databases and repositories. However, given the different licences and interpretations available, OpenScienceLink must continuously verify whether its use falls within the scope of the terms and conditions under which third party content has been made available to the public.

*Next steps developments of the licensing platform*

- OpenScienceLink platform cannot achieve its goals of open access to scientific data unless the proposed platform services are either authorized by the respective rights owners or benefit from a legal exception. However, as was discussed in the previous sections, the scope of exceptions significantly differs among Member States and does moreover not apply to all OpenScienceLink platform services. Further guidelines for open access licensing will be proposed in work package 8.

[340] The license framework will be further developed and evaluated based on the user feedback and changes of the OpenScienceLink Platform services. Special attention will be given to the possibilities to allow for less restrictive licensing and license interoperability to allow for a more flexible and open platform where research data can be shared.

### 8.1.4 LIABILITY

- Having assessed the scope of the E-Commerce Directive, OpenScienceLink, caution has been called for as it is unlikely that OpenScienceLink will be able to benefit from an exemption of liability as a hosting provider.
- For services that do not fall under the liability exemptions, minimizing the risk of infringement of content being used and made available on the platform should be a priority when further developing the platform services. An assessment of the tools and services that are available is necessary to see what would best fit within the OpenScienceLink platform.
- Although there is no general obligation to monitor the content of the platform, there is a possibility that member states require ISPs to report illegal activity or take down illegal information.
- Considering the above, it will be important for the OpenScienceLink platform to have in place an efficient 'notice and take down' procedure as well as relevant provisions regarding the limitations of its liability. Best practices will be assessed to see what procedure would best fit the OpenScienceLink platform.